Establishing sequence and structural requirements for human Y5 RNA cleavage

Martina Billmeier

A thesis submitted for the degree of
Doctor of Philosophy

University of East Anglia
School of Biological Sciences

October 2019

Word count: 78345

“In all things of nature there is something of the marvellous”

Aristotle (384-322 BC)

Abstract

Small non-coding RNAs (sRNAs) are important regulators of gene expression and play fundamental roles in many different organisms. During the last decade, next generation sequencing (NGS) has become a powerful tool for sRNA profiling and several classes of novel regulatory sRNAs with potential biological functions have been discovered. Among them, sRNAs derived from longer non-coding Y RNAs were identified.

Y RNAs are non-coding RNAs (~ 100 nt) that are evolutionarily conserved in vertebrates. Human Y RNAs bind to the autoimmune proteins Ro60 and La forming ribonucleoprotein complexes. Y RNAs are involved in the initiation of chromosomal DNA replication, regulation of RNA stability and cellular stress response. During apoptosis, Y RNAs produce smaller RNA fragments from the 3' and 5' ends. However, the biogenesis, processing and the biochemical functions of Y RNA derived sRNAs remain elusive.

In previous work a high throughput mutagenesis analysis on the 3' end of hY5 RNA was performed. In this work, I confirmed that Y RNA cleavage on the 3' end of hY5 RNA correlated with the secondary structure of the mutated hY5 RNA. Furthermore, I clearly showed that the nucleotide sequence in the loop was not important and a bulge of one nucleotide was sufficient for Y RNA cleavage to occur.

In this thesis I designed and performed a high throughput mutagenesis approach on the 5' end of hY5 RNA. This revealed a nucleotide motif UUAU located in an internal loop at the position 22-25 which contributed to Y RNA cleavage at both the 3' and 5' end of hY5 RNA.

In order to identify ribonucleases involved in Y RNA cleavage, I confirmed that Ro60 was essential for YsRNA production in mouse embryonic stem cells. Interestingly, I could show that RNase L contributed to Y RNA cleavage in mouse embryonic fibroblast cells and human lung cancer cells.

Table of contents

Abstract **3**
Table of contents **4**
List of figures **9**
List of tables **11**
Acknowledgements **12**
List of abbreviations **14**

Chapter 1 Introduction **20**
1.1 The central dogma of molecular biology 20
1.2 The non-coding RNA world 20
1.2.1 Ribosomal RNAs and transfer RNAs 20
1.2.2 Small nuclear RNAs 21
1.2.3 Small nucleolar RNAs 21
1.2.4 Vault RNAs 22
1.3 Small non coding RNAs 23
1.3.1 Discovery of small non-coding RNAs 23
1.3.2 MiRNAs 24
1.4 Small RNAs derived from longer transcripts 26
1.4.1 SnoRNA derived small RNAs 26
1.4.2 Transfer RNA derived small RNAs 26
1.4.3 Vault RNA derived small RNAs 29
1.5 Y RNAs 30
1.5.1 Y RNAs in general 30
1.5.2 Evolution of Y RNAs 31
1.5.3 Secondary structure of Y RNAs 34
1.6 Y RNAs in RoRNPs 40
1.6.1 Ro60 protein 40
1.6.2 Role of RoRNPs in RNA quality control 41
1.6.3 Role of RoRNPs in stress response 42
1.6.4 Subcellular localisation of RoRNPs 43
1.7 Y RNA binding proteins 44
1.7.1 La protein 44
1.7.2 Nucleolin 45
1.7.3 Ro52 46

1.7.4 Heterogeneous nuclear ribonucleoproteins 47
1.7.5 Zipcode-binding protein 47
1.7.6 APOBEC3G 48
1.7.7 RoRNP binding protein 1 48
1.7.8 Ribosomal protein L5 49
1.7.9 Poly (A) specific ribonuclease 50
1.7.10 Cleavage and polyadenylation specificity factor 50
1.8 Functions of Y RNAs 51
1.8.1 Y RNAs in DNA replication 51
1.8.2 Y RNAs in cancer 57
1.9 Y RNA derived small RNAs 58
1.10 Functions of Y RNA derived small RNAs 60
1.10.1 Y RNA derived small RNAs involved in coronary artery disease 60
1.10.2 Y RNA fragment involved in cardioprotection 61
1.10.3 Y RNA derived sRNAs regulate cell death and inflammation in monocytes/macrophages 61
1.10.4 Human Y4 RNA fragment can act as a guide for tRNase Z 63
1.10.5 PTBP1 63
1.10.6 RNase L 64
1.11 Aims and objectives of the thesis 66

Chapter 2 Materials and Methods 70

2.1 Cell culture 70
2.1.1 Cell lines and cell culture media 70
2.1.2 Sub-culturing and cell culture conditions 71
2.1.3 Freezing cells 71
2.1.4 Transfection of NIH 3T3 cells with plasmid DNA 72
2.1.5 Transfection of MCF7 cells and mES wild type and Ro60 $^{-/-}$ cells with poly (I:C) 72
2.1.6 Treatment of NIH 3T3 cells with staurosporine 73
2.2 Protein extraction 73
2.3 RNA extraction 74
2.4 Detecting small RNAs by Northern blotting 75
2.4.1 Separation of small RNAs by denaturing polyacrylamide gel electrophoresis 75
2.4.2 Transfer of small RNAs onto a nylon membrane using semi-dry transfer 76
2.4.3 Chemical cross-linking 76

2.4.4 Hybridization 77
2.4.5 Washing of the membrane and signal detection 78
2.4.6 Stripping of Northern blot membranes 78
2.4.7 Oligonucleotide probe sequences used for Northern blotting 78
2.5 Western blotting 79
2.5.1 Polyacrylamide gel electrophoresis of proteins 79
2.5.2 Western blot transfer 79
2.5.3 Blocking of membrane and antibody hybridization 79
2.5.4 Western blot antibodies 80
2.6 Generation of DNA constructs 80
2.6.1 DNA extraction from HeLa cells 80
2.6.2 Amplification of the human Y5 RNA gene from genomic DNA 80
2.6.3 Agarose gel electrophoresis 81
2.6.4 Purification of PCR products by gel extraction 81
2.6.5 Adenosine tailing of PCR products for TA cloning 82
2.6.6 Cloning of PCR products into pGEMT easy vector 82
2.6.7 Preparation of supercompetent DH5α bacterial cells 84
2.6.8 Transformation of bacterial cells 84
2.6.9 Colony PCR 85
2.6.10 Purification of plasmid DNA 86
2.6.11 Sanger sequencing 86

2.7 High throughput mutagenesis approach on the 3' end of the human Y5 RNA 86

2.7.1 Generation of the 3' end hY5 RNA mutant pool small RNA cDNA libraries 86
2.7.2 Generation of the 3' end hY5 RNA mutant pool full length cDNA libraries and plasmid libraries 87
2.7.3 Bioinformatics analysis of hY5 3' end mutant sRNA libraries 88
2.7.4 Bioinformatics analysis of the full-length cDNA and plasmid libraries 89
2.7.5 Secondary structure predictions and visualization 89
2.7.6 Generation of individual hY5 3' end mutants and loop mutants 89
2.7.7 Transfection of hY5 mutants into NIH 3T3 cells 90
2.7.8 Treatment of NIH 3T3 cells with staurosporine 90
2.7.9 RNA extraction and Northern blot analysis 91

2.8 High throughput mutagenesis approach on the 5' end of the human Y5 RNA 91

2.8.1 Generation of 5' end hY5 mutant pools-first strategy..... 91
2.8.2 Generation of hY5 5' end mutant pool plasmid libraries 93
2.8.3 Generation of the hY5 5' end mutant pools- second cloning strategy..... 95
2.8.4 Generation of full length cDNA and plasmid libraries for the hY5 5' end mutant pools 96
2.8.5 Small RNA library construction of hY5 5' mutants 97
2.8.6 Generation of individual hY5 5' end mutants 100
2.9 Small RNA library construction using High Definition adapters..... 100

Chapter 3 Establishing sequence and structural requirements for 3' end human Y5 RNA cleavage 101

3.1 Introduction 101
3.1.1 Cellular stress response 101
3.1.2 Previous work and background of the project..... 102
3.2 Results..... 106
3.2.1 Bioinformatics analysis of 3' hY5 mutagenesis libraries 106
3.2.2 Some library 1 hY5 RNA mutants generate longer YsRNAs compared to hY5 mutant pool 2 and 3 113
3.2.3 Library 2 hY5 RNA mutants generate YsRNAs at the same size as wild type..... 130
3.2.4 Library 3 hY5 RNA mutants generate YsRNAs at the same size as wild type hY5 RNA 137
3.2.5 A bulge of one nucleotide is required for human Y5 RNA cleavage 144
3.2.6 The sequence of the L2a loop does not seem to be important for human hY5 RNA cleavage..... 148
3.3 Discussion 157

Chapter 4 Establishing sequence and structural requirements for 5' end human Y5 RNA cleavage 162

4.1 Introduction 162
4.1.1 Extracellular vesicles 162
4.1.2 Functions of 5' end derived Y RNAs in extracellular vesicles 163
4.2 Results 165
4.2.1 Generation of 5' end hY5 mutant plasmid pools 165
4.2.2 Generating 5' end hY5 mutant plasmid pools 171
4.2.3 Bioinformatics analysis of the 5' end hY5 mutagenesis libraries 176
4.2.4 Library 4 hY5 RNA mutants generate wild type sized YsRNAs 185
4.2.5 Library 5 hY5 RNA mutants produce wild type sized YsRNAs 190
4.2.6 Library 6 hY5 RNA mutants generate YsRNAs at the same size as wild type 193
4.2.7 Is the loop L2b important for human Y5 RNA cleavage? 196
4.3 Discussion 211

Chapter 5 Investigation of proteins involved in Y RNA cleavage 216

5.1 Introduction 216
5.1.1 Role of RNA binding proteins in cellular stress conditions 216
5.1.2 Y RNAs as part of RoRNPs 217
5.1.3 Role of ribonucleases during stress 217
5.1.4 Previous work and aims of this chapter 218
5.2 Results 219
5.2.1 Ro60 is required for Y RNA cleavage 219
5.2.2 Are there other Ro60 dependent small RNAs? 231
5.2.3 Is RNase L involved in Y RNA cleavage? 237
5.3 Discussion 243

Chapter 6 Summary and general discussion 250

References 258

Appendices 279

List of figures

Chapter 1:

Figure 1.1. Secondary structure predictions of hY5 RNA using Mfold and chemical structure probing experiments. 36
Figure 1.2. Secondary structures and modular composition of eukaryotic and prokaryotic Y RNAs. 38
Figure 1.3. Overview of DNA replication. 53
Figure 1.4. Overview of aims and objectives performed in this thesis. 69

Chapter 2:

Figure 2.1. Vector map of pGEMTeasy and pGEM-hY5 WT plasmid DNA vector. . 83

Chapter 3:

Figure 3.1. Overview of the design and workflow of the 3' end high throughput mutagenesis on the 3' end of the hY5 RNA. 104
Figure 3.2. Requirement of a single cytosine bulge and a G-C base pair between the positions 8 and 68 for Ro60 binding and hY5 RNA stability. 111
Figure 3.3. Some library 1 hY5 RNA mutants generate longer YsRNAs compared to mutant pool 2 and 3. 116
Figure 3.4. Secondary structure features of library 1 hY5 RNA mutants correlate with hY5 RNA cleavage. 119
Figure 3.5. Most abundant library 1 hY5 RNA mutants generate longer YsRNAs. 124
Figure 3.6. Library 1 hY5 mutants with exactly the same secondary structure as wild type produce wild type sized YsRNAs compared to the most abundant hY5 library 1 mutants that produce longer YsRNAs. 129
Figure 3.7. Library 2 hY5 RNA mutants produce YsRNAs at the same size as wild type. 135
Figure 3.8. Library 3 hY5 RNA mutants generate YsRNAs at the same size as wild type hY5 RNA. 142
Figure 3.9. A bulge of one nucleotide is required for hY5 RNA cleavage. 147
Figure 3.10. The cytosines at the positions 50, 51 and 52 are not essential for hY5 RNA cleavage 154
Figure 3.11. Substitution mutations of hY5 RNA at position 49 did not have any effect on hY5 RNA cleavage. 156

Chapter 4:

Figure 4.1. Design and generation of the 5' end hY5 high throughput mutagenesis approach. 170
Figure 4.2. First cloning strategy for 5' end high throughput mutant pool generation. 172
Figure 4.3. Second cloning strategy for 5' end high throughput mutant pool generation. 175
Figure 4.4. Plasmid libraries of hY5 libraries 4, 5 and 6. 179
Figure 4.5. Full length cDNA libraries of hY5 library mutant pool 4, 5 and 6. 181
Figure 4.6. Generation of 5' end derived YsRNA cDNA libraries. 184
Figure 4.7. Library 4 hY5 RNA mutants generate wild type-sized YsRNAs. 189

Figure 4.8. hY5 library 5 mutant Y RNAs produce wild type sized YsRNAs. 192
Figure 4.9. Library 6 mutant Y RNAs produce wild type sized YsRNAs..... 195
Figure 4.10. The uridine at position 22 is essential for 5' end hY5 RNA cleavage and might be involved in 3' end hY5 RNA cleavage. 199
Figure 4.11. The internal loop L2b seems to be important for human Y5 RNA cleavage from both the 3' and 5' end of human Y5 RNA. 206
Figure 4.12. The wild type sequence motif UUAU of the internal loop L2b is important for 3' and 5' end hY5 RNA cleavage. 210

Chapter 5:

Figure 5.1. Ro60 contributes to Y RNA cleavage in mES cells..... 221
Figure 5.2. Generation of small RNA cDNA libraries of wild type and Ro60 $^{-/-}$ mES cells treated with poly (I:C) or mock treated using High definition adapter sequences. 225
Figure 5.3. Ro60 is required for Y RNA cleavage..... 230
Figure 5.4. Ro60 seems not be involved in other Ro60 dependent sRNA generation mechanisms. 236
Figure 5.5. RNase L contributes to Y RNA cleavage in mEF cells and human A549 lung cancer cells upon poly (I:C) treatment but is not involved in 3' end hY5 RNA cleavage..... 241
Figure 5.6. Model of the role of Ro60 in YsRNA generation and protection of Y RNAs from degradation. 246

Chapter 6:

Figure 6.1. Graphical summary of the sequence and structural features identified from the 3' and 5' end high throughput mutagenesis analysis..... 255

List of tables

Chapter 3:

Table 3.1. Overview of sequenced 3' end hY5 plasmid bias libraries. 108
Table 3.2. Sequencing information of 3' end hY5 full length cDNA libraries. 108

Appendix II:

Table 0.1. List of DNA oligonucleotides used for hY5 gene amplification. 279
Table 0.2. DNA oligonucleotides used for colony PCR. 279
Table 0.3. DNA oligonucleotides used for generating the 3' end hY5 RNA mutants. 279
Table 0.4. DNA oligonucleotides used for generating "balancing" DNA. 283
Table 0.5. DNA oligonucleotides used to generate the hY5 mutant pool PCR products in strategy 1. 283
Table 0.6. DNA oligonucleotides used for generating 5' end hY5 plasmid libraries. 284
Table 0.7. DNA oligonucleotides used to generate the hY5 mutant pool PCR products in strategy 2. 285
Table 0.8. DNA oligonucleotides used for the final overlap extension PCR. 285
Table 0.9. DNA oligonucleotides used for generating full length cDNA libraries ... 286
Table 0.10. DNA oligonucleotides used for small RNA library construction of 5' end derived YsRNA mutants. 286
Table 0.11. DNA oligonucleotides used for generating individual 5' end derived hY5 RNA mutant sequences. 287
Table 0.12. List of DNA oligonucleotides used for Northern blotting. 289
Table 0.13. Sequencing information of 3' end hY5 small RNA cDNA libraries. 290
Table 0.14. General sequencing information of wild type and Ro60 $^{-/-}$ mES sRNA cDNA libraries either treated or untreated with poly (I:C). 291
Table 0.15. Differential gene expression analysis of mY1 and mY3 in wild type and Ro60 $^{-/-}$ cells untreated or treated with poly (I:C) using DESeq2. 292

Acknowledgements

First of all, I would like to thank my supervisor and group leader Prof. Tamas Dalmay for everything he did for me since I came to his laboratory five years ago during my ERASMUS internship. I am very grateful that he always believed in me and that he gave me the opportunity to pursue my PhD project in his research group. His great expertise in RNA biology, inspiring scientific discussions, optimism and patience always encouraged me throughout my PhD.

My special thanks goes to Ping Xu for supervising me during my master's internship and two years of my PhD. Without her knowledge, scientific expertise and never-ending discussions about science and life in general I would not be there where I am now. I am extremely grateful for all her advice and everything I learnt from her as a scientist and a person. I really appreciate all her attention and effort she put in me that I became an independent researcher. Also thank you very much for being my friend and always encouraging me also in difficult times when I came across some troubles.

I am indebted to Dr. Simon Moxon for being my secondary supervisor and for the huge amount of bioinformatics analysis he did for me in the last years. I really appreciate his valuable suggestions and all the help throughout these years.

I very thankful to all the colleagues and students I was working with in the last years here at the University of East Anglia. Throughout these years I have met so many people from many different countries and cultures and I have been every lucky to able to work with all of you.

My particular thanks go to my best Indian friend, colleague and now postdoctoral researcher Dr. Archana Singh for all her loving care, help in bioinformatics and fun times we always had and we will still have. For sure, without all your food, bollywood movies and taking care of me I would not have survived the PhD. I will always remember our extreme camping trip to Snowdonia. For sure I will miss all the weekends with Kofra coffee and gossip. Archana, with all your energy and strength you brightened up every day of my PhD life and I will miss you the most!

Thank you very much also to Aleks for all his care and support as well as the exciting trips and movie nights.

I would also like to thank my previous lab members and friends Dr. Aurore Coince, Dr. Zahara Medina Calzada and Dr. Afroditi Tsaballa for all her help, advice and fun times outside of the lab.

My special thanks goes to Maria-Elena Mannarelli for all her help and valuable discussions in the lab or life in general. In particular, I would like to thank her and Tim for providing me an accommodation in the last year.

Besides, thank you very much to my previous lab members Dr. David Collins, Dr. Dave Kushner, Dr. Trung Trieu for their advice and helpful discussions.

From the lab I would like to thank Dr. Darrell Green for his friendship and expertise. I am also grateful to Rocky Payet, Yvonne Ridge and Mamdouh Alshammari for their help in the daily research life. I would like to thank Afrah and Lerato for all the nice and encouraging conversations in the lab and it was a pleasure teaching you ☺

I would also like to thank my other friends from Norwich I met during the last few years. Thank you to Athina, Alaa, Sue, Faye, and Qamar for being the best housemates in Caddow Road. I am extremely grateful to Athina for her friendship, especially all her care and help when I came to Norwich for the first time. Thank you very much also to Isabella for all the nice conversations and coffee breaks we had during the last years.

Finally, I would like to thank my family for all their love, support and faith in me. Thank you very much for always being there when I needed you!

List of abbreviations

A	
A	adenine
A. thaliana	*Arabidopsis thaliana*
APS	Ammonium persulfate
APOBEC3G	Apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G
ATP	Adenosine triphosphate
ASR	AGO-taxis small RNAs
B	
bp	base pair
C	
C	cytosine
CAD	coronary artery disease
cDNA	complementary DNA
CDC	cardiosphere derived cells
C. elegans	*Caenorhabditis elegans*
CD	circular dichroism
CDS	coding sequence
CMCT	1-cyclohexyl-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate
CPSF	cleavage and polyadenylation specificity factor
CRISPR	Clustered regularly interspaced short palindromic repeats
D	
DGCR8	DiGeorge Syndrome Critical Region 8
DNA	Deoxyribonucleic acid
DMEM	Dulbecco's modified eagle medium
DMS	Dimethylsulfide
DMSO	Dimethylsulfoxide
dNTPs	desoxyribonucleotides
ds	double-stranded

dsRNA	double-stranded RNA
DTT	Dithiothreitol
E	
EBER	Epstein-Barr virus-encoded small RNA
EBV	Epstein-Barr virus
EDC	Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride
EDTA	Ethylenediaminetetraacetic acid
E.coli	*Escherichia coli*
EV	Extracellular vesicle
F	
FBS	Fetal bovine serum
FL	Full length
G	
G	guanine
gDNA	genomic DNA
Gln	Glutamine
G-phase	Gap phase of cell cycle
H	
h	hour
HD	High definition adapter
HF	High fidelity
hnRNP	heterogeneous nuclear ribonucleoprotein
His	Histidine
hY5 RNA	human Y5 RNA
I	
IFIT5	Interferon-induced protein with tetratricopeptide repeats 5
IGV	Integrative genome browser
IL	interleukin
itRFs	internal tRNA fragments
IPTG	Isopropyl -D-1-thiogalactopyranoside

J

K

kb	kilobase
KH domain	hnRNP K homology
KD	Knockdown
KO	Knockout

L

L	Library
L	Loop (secondary structure)
LB	Lysogenic broth
LNA	Locked nucleic acid

M

M	Mutant
MBq	Mega Becquerel
MBT	Midblastula transition
Mcm	minichromosome maintenance protein complex
mEF	mouse fibroblast cells
mES	mouse embryonic stem cells
$MgCl_2$	magnesium chloride
MIDAS	metal ion dependent adhesion site
miRNA	micro RNA
min	minutes
MOV	Moloney leukaemia virus
mRNA	messenger RNA
MVB	Multivesicular bodies
mY1	mouse Y1 RNA
mY3	mouse Y3 RNA

N

N	random nucleotides
ncRNA	Non-coding RNA
NGS	Next generation sequencing
NMR	nuclear magnetic resonance

nt	nucleotide
O	
OAS	2’ 5’ Oligoadenylate pathway
ORC	Origin recognition complex
P	
PA	palmitic acid
PAGE	Polyacrylamide gel electrophoresis
PARN	Poly(A) specific ribonuclease
PAZ	Piwi/Argonaute/Zwille
PB	Plasmid bias
PBS	Phosphate buffered saline
PCA	Principle component analysis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
piRNA	Piwi-interacting RNAs
PKR	Protein kinase RNA-activated
PNK	T4 polynucleotide kinase
PNPase	Polynucleotide phosphorylase
PPR	pathogen pattern recognition receptor
pppG	triphosphorylated guanosine residue
pppA	triphosphorylated adenosine residue
pre-RC	prereplicative complex
PTBP1	Polypyrimidine tract-binding protein
PUF	poly (U) splicing factor
Pol	Polymerase
Poly (I:C)	Polyinosinic: polycytidylic acid
Pro	Proline
Q	
qPCR	Quantitative PCR
R	
RBP	RNA binding protein

RISC	RNA-induced silencing complex
RoBPI	Ro RNP binding protein
ROP-1	Ro60 homologue in *Caenorhabditis elegans*
Ro RNP	Ro60 ribonucleoprotein
RNA	ribonucleic acid
RNAi	RNA interference
RNase	ribonuclease
RNP	ribonucleoprotein
RPA	Replication protein A
RPC	Replication protein C
RPM	revolutions per minute (centrifuge)
RPM	reads per million (bioinformatics)
rRNA	ribosomal RNA
Rsr	Ro60 related protein
RRM	RNA recognition motif
RT-PCR	Reverse transcriptase PCR
RYPER	Ro60/Y RNA/PNPase Exoribonuclease RNP
S	
s	second
sbRNA	stem bulge RNA
SDS	Sodium dodecyl sulfate
siRNA	small interfering RNA
SLE	Systemic lupus erythematosus
snoRNA	small nucleolar RNA
snRNA	small nuclear RNA
S-phase	Synthesis phase of cell cycle
sRNA	small RNA
ss	single-stranded
SS	Sjögren's syndrome
SSC	Saline-Sodium citrate
stRNA	small temporal RNA
STS	Staurosporine

svRNA	small vault-derived RNA
T	
T	thymine
TBE	Tris-borate-EDTA
TEMED	Tetramethylethylenediamine
tRF	transfer RNA regulatory fragment
TLR	Toll like receptor
tRNA	transfer RNA
U	
U	uridine
UTR	Untranslated region
UV	Ultraviolet light
U RNA	small non-coding nuclear RNA
V	
V	voltage
vRNA	vault RNA
vWFA	von Willebrand factor A-like motif
W	
WT	wild type
X	
X-gal	5-Bromo-4-chloro-3-indolyl -D-galactopyranoside
X. laevis	*Xenopus laevis*
Y	
YB-1	Y box-binding protein 1
Y RNA	cytoplasmic RNA (Y1, Y3, Y4, Y5)
Y RNP	Y RNA ribonucleoprotein
YsRNA	Y RNA derived small RNA
Z	
ZBP1	Zipcode-binding protein 1

Chapter 1 Introduction

1.1 The central dogma of molecular biology

The central dogma is a fundamental concept of molecular biology that was initially proposed by Francis Crick in 1958. It describes the flow of genetic information from DNA via RNA to proteins (Crick, 1958). In 1970 the central dogma had to be redefined due to the discovery of retroviruses that contain RNA genomes and generate DNA by reverse transcription. The restated dogma emphasized that genetic information cannot be transferred from proteins to other proteins or nucleic acids (Crick, 1970).

Intriguingly, the human genome project revealed that the majority of the human genome is comprised of non-coding DNA regions and only 1.5% of the human genome encode for proteins. It was shown that around 98% of the human transcriptome consists of non-protein coding RNA. Initially it was thought that these non-coding regions are “junk DNA” and not functional (Mattick, 2001; Venter et al., 2001; International Human Genome Sequencing Consortium, 2004; Birney et al., 2007).

However, with advances of sequencing technologies in the last decade it became apparent that intronic and other non-protein coding RNAs fulfil vital functions (Mattick, 2004). Consequently, the central dogma has been challenged because these regulatory non-coding RNAs represent an additional layer of gene regulation (Mattick, 2003). Non coding RNAs play key roles in RNA metabolism and are involved in almost all cellular processes including development, differentiation, stress response, epigenetics, maintaining genome stability and diseases (Morris and Mattick, 2014).

1.2 The non-coding RNA world

1.2.1 Ribosomal RNAs and transfer RNAs

Ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs) were initially discovered in the 1950s and have been intensively studied due to their high abundance in cells. The majority of total RNA in cells is comprised of rRNAs (80-90%) and tRNAs (10-12%) whereas messenger RNAs (mRNAs) only make up a small proportion (3-7%) and all other non-coding RNAs together constitute less than 2% of the total cellular RNA mass (Palazzo and Lee, 2015).

Ribosomal RNAs were discovered by electron microscopy as part of the ribosome which is essential for protein synthesis (Palade, 1955). In 1958 transfer RNAs were described as adaptor molecules and mediators between mRNA and amino acid sequence of proteins during translation (Crick, 1958; Hoagland et al., 1958).

1.2.2 Small nuclear RNAs

Apart from rRNAs and tRNAs that are infrastructural components of protein synthesis small nuclear RNAs (snRNAs) were among the first discovered non-coding RNAs. Small nuclear RNAs were identified as components of ribonucleoprotein particles in patients with the autoimmune disorder systemic lupus erythematosus (Lerner and Steitz, 1979). These ncRNAs included U1, U2, U4, U5 and U6 snRNAs and have been shown to be part of the spliceosome machinery and mRNA splicing (Kruger et al., 1982; Krämer, 1996).

1.2.3 Small nucleolar RNAs

Small nucleolar RNAs (snoRNAs) are another important class of non-coding RNAs that were described in the 1990s (Maxwell and Fournier, 1995; Balakin et al., 1996; Weinstein and Steitz, 1999). SnoRNAs are ~ 70-250 nt in length, are highly abundant in the nucleolus and Cajal bodies of eukaryotic cells and have been shown to guide chemical modifications and processing of several RNA species including rRNAs, snRNAs and tRNAs. SnoRNAs can be classified based on their structural features and chemical RNA modification they mediate into C/D box snoRNAs and H/ACA box snoRNAs. C/D box snoRNAs are responsible for the 2’ O-methylation of rRNAs, snRNAs and tRNAs whereas H/ACA box snoRNAs direct pseudoruridylation of RNAs (Kiss, 2001, 2002). SnoRNAs have also been shown to be involved in alternative splicing (Kishore and Stamm, 2006; Falaleeva et al., 2016), chromatin remodelling (Leung et al., 2009; Schubert et al., 2012) and diseases such as cancer (Mannoor et al., 2012).

1.2.4 Vault RNAs

Vault RNAs (vtRNAs) are part of a 13 MDa vault ribonucleoprotein complex and range in sizes between 80-150 nts (Kedersha and Rome, 1986; Stadler et al., 2009). In humans there are four paralogues described (vtRNA1-1, 1–2, 1–3 and 2–1). Vault ribonucleoprotein complexes RNAs are transcribed by RNA polymerase III, are highly conserved among most eukaryotes and are characterized by a hollow barrel-shaped structure. The function of vault RNAs is still not very clear but evidence suggested that they play a role in chemotherapy resistance of cancer cells (Kickhoefer et al., 1998; Steiner et al., 2006), nuclear transport (Vollmar et al., 2009), immunity (Berger et al., 2008; Nandy et al., 2009), signalling, (Kim et al., 2006; Steiner et al., 2006; Berger et al., 2008), DNA damage repair (Shimamoto et al., 2006) and apoptosis (Amort et al., 2015). Recently, vault RNAs have been reported to regulate autophagy (Horos et al., 2017).

In the last decades many other novel classes of non-coding RNAs have been identified due to the further advances and improvement of next generation sequencing technologies. Non coding RNAs represent key molecules that define and contribute to the eukaryotic complexity (Mattick, 2001). The discovery of non-coding RNAs has revolutionized the RNA world and adds an additional layer of regulation and complexity (Mattick, 2003; Hall and Dalmay, 2013).

1.3 Small non coding RNAs

Historically, the term small RNA was used to describe all RNA molecules between 100-200 nt in prokaryotes (Wassarman et al., 1999; Storz, 2002).

However, due to the discovery of RNA interference (RNAi) and small non-coding RNAs in plants and animals in the 1990s and 2000s all RNA molecules less than 200 nucleotides (nt) were named small non-coding RNAs. More recently the term small RNA (sRNA) was used for the really small RNA molecules, i.e. 20-30 nt (Rother and Meister, 2011).

In contrast, long non-coding RNAs are usually longer than 1 kilobase (1 kb) up to several thousands of nt (Storz, 2002; Rother and Meister, 2011).

1.3.1 Discovery of small non-coding RNAs

The first small non-coding RNAs lin-4 and let-7 (~22 nt) were discovered in the nematode *Caenorhabditis elegans* (*C. elegans*) through genetic mutant screening (Lee et al., 1993; Reinhart et al., 2000). These RNAs were expressed in a developmental stage-specific manner and were termed as short temporal RNAs (stRNAs). They were shown to control the timing of transition during the development of *C. elegans* by binding to complementary sequences in the 3' untranslated regions (3'-UTR) of development-specific mRNAs resulting in the inhibition of gene expression (Pasquinelli et al., 2000).

In 1998, Fire and Mello found in *C. elegans* that long double stranded (ds) RNA results in the sequence-specific mRNA degradation and inhibition of gene expression through a phenomenon called RNA interference (RNAi) (Fire et al., 1998). Later RNAi turned out to be the same process as what was called post-transcriptional gene silencing in plants (Jorgensen, 1990; Napoli et al., 1990).

One year later Hamilton and Baulcombe discovered an abundant class of small dsRNAs (21-24 nt) that were correlated with post-transcriptional gene silencing (Hamilton and Baulcombe, 1999). Further biochemical studies on *Drosophila* cell extracts demonstrated that long dsRNA molecules were processed into small interfering RNAs (siRNAs) by endonucleolytic cleavage through the RNase III enzyme Dicer in both animals and plants (Hammond et al., 2000; Zamore et al., 2000; Bernstein et al., 2001; Elbashir et al., 2001). Interestingly, stRNAs had the same size

as siRNAs and the let-7 gene was shown to be conserved across the animal kingdom. Therefore, it was hypothesized that Dicer might produce stRNAs which would mean that they are not unique to *C. elegans* and that there could be more stRNAs in different species. These hypotheses were tested by looking for 20-24 nt small RNAs in several species including *C.elegans*, *Drosophila* and the human HeLa cell line.

In 2001, three laboratories identified over one hundred new sRNAs (~21-25 nt) in *Drosophila* embryos, *C. elegans* and human HeLa cells by cloning size-fractionated sRNAs and sequencing the cDNA libraries. Some of these newly identified sRNAs were similar in size to the previously discovered stRNAs lin-4 and let-7. These RNAs were encoded by the genome, were expressed in specific tissue- and/or at developmental stages and were conserved among species. Interestingly, a few of these identified sRNAs were constitutively expressed independent of developmental stage or tissue. Therefore, this novel class of small non-coding RNAs was termed microRNAs (miRNAs) instead of stRNAs (Lau et al., 2001).

1.3.2 MiRNAs

The intensive cloning of small RNAs (sRNAs), the improved sequencing of cDNA libraries and bioinformatics analysis revealed hundreds of miRNA genes in animals, plants, and viruses (Bartel, 2004; Pfeffer et al., 2004). Data from the miRNA database (miRBase 21) revealed that there are 2588 predicted miRNAs in human (Griffiths-Jones, 2006; Ha and Kim, 2014).

MiRNAs constitute one of the most abundant and best studied classes of small non-coding RNAs in human. Intriguingly, around 30% of mammalian protein-coding genes are regulated by miRNAs (Filipowicz et al., 2008). MiRNAs are around 19-23 nt in length and regulate gene expression by binding to partially complementary (animals) or near perfectly complementary sequences (plants) of target mRNAs which results in translational repression and degradation of mRNAs (Huntzinger and Izaurralde, 2011).

The majority of miRNA genes are transcribed by RNA polymerase II from intergenic regions or from introns of protein coding genes. Primary miRNA transcripts (pri-miRNAs) are cleaved in the nucleus by the microprocessor complex which includes the RNase III enzyme Drosha, the dsRNA-binding domain protein DiGeorge syndrome critical region 8 (DGCR8) and further auxiliary proteins resulting in the

generation of precursor miRNAs (pre-miRNAs) (Denli et al., 2004; Han et al., 2004). These pre-miRNAs are transported from the nucleus via the nuclear core complex into the cytoplasm by Exportin-5 via a Ran-GTP dependent mechanism (Bohnsack et al., 2004). In the cytoplasm the RNase III enzyme Dicer cleaves the pre-miRNAs into short ~ 22 nt miRNA duplex intermediates with 5' phosphates and 2 nt 3' overhangs (Hutvágner et al., 2001; Ketting et al., 2001). The miRNA duplex structures are comprised of a guide strand and passenger strand (miRNA/miRNA*) (Kim et al., 2009).

During assembly of the RNA induced silencing complex (RISC) the miRNA duplex structure gets unwound and only one strand is incorporated into the RISC complex, whereas the other strand is usually removed and degraded. However, there are reports that the miRNA*sequences can be loaded into RISC and act as functional miRNAs (Okamura et al., 2008; Ghildiyal et al., 2010).

The determination which strand gets incorporated and retained in the RISC complex correlates with its thermodynamic stability and is mainly based on the asymmetry rule. The strand with the lower thermodynamic stability at the 5' end will be incorporated into the RISC complex and guides the complex to the mRNA target (Khvorova et al., 2003; Schwarz et al., 2003; Okamura et al., 2008).

In the RISC complex the mature miRNA binds to a member of the Argonaute protein family.

Argonaute proteins are characterized by four domains, N-, PIWI Argonaute Zwille (PAZ), middle (MID) - and PIWI domains. The N-domain is involved in loading of the sRNAs into the RISC complex and functions in the unwinding of the miRNA duplex during the RISC assembly process (Kwak and Tomari, 2012). The PAZ domain interacts with the 3' end of sRNAs and the MID domain recognizes the 5' end of sRNAs (Hutvagner and Simard, 2008). The PIWI domain has structural similarities to RNase H, which cleaves RNA in DNA-RNA hybrids and was shown to have endonuclease activity (Rivas et al., 2005).

In human only Ago2, but not Ago1, Ago3 and Ago4 have been shown to have endonucleolytic activity (Meister, 2013). After loading into the RISC complex, posttranscriptional gene silencing is mediated through binding of miRNAs to partially complementary target sites within the 3' untranslated region (3' UTR) of their target mRNAs resulting in translational repression and/or mRNA destabilization of mRNAs (Filipowicz et al., 2008; Bartel, 2009).

During the last decade several novel classes of sRNAs apart from miRNAs have been discovered due to advancing next generation sequencing technologies. Interestingly,

many small RNA fragments derived from longer non-coding RNAs have been identified and it was shown that some of these sRNAs are functional and not degradation products (Wittmann and Jäck, 2010; Rother and Meister, 2011). In recent years it has been shown that small RNA fragments can be generated from several classes of non-coding RNAs, including snoRNAs, snRNAs, rRNAs, tRNAs, vault RNAs or Y RNAs (Li et al., 2012; Chen and Heard, 2013).

In the next section I will briefly give an overview to the most intensively studied sRNA fragments that are generated from non-coding RNAs.

1.4 Small RNAs derived from longer transcripts

1.4.1 SnoRNA derived small RNAs

Small RNAs derived from snoRNAs (sdRNAs) have been initially discovered by next generation sequencing of immunoprecipitated RNAs associated with Ago1 and Ago2 using HEK293 cells. Among the sequencing reads, the sdRNA ACA45 was identified. Interestingly, this sRNA was dependent on Dicer, but independent of Drosha/DGCR8. Luciferase experiments revealed that this sdRNA mediated post transcriptional gene silencing and was acting such as miRNAs (Ender et al., 2008). In animals, sdRNAs derived from H/ACA box snoRNAs are mainly 20-24 nt and are generated from the 3' end. C/D box sdRNAs are 17-19 nt or longer than 27 nt and are mainly derived from the 5' end of snoRNAs (Taft et al., 2009; Brameier et al., 2011). In the last decade several sdRNAs with miRNA-like functions have been identified. They have been shown to be involved in alternative splicing (Kishore and Stamm, 2006; Kishore et al., 2010) and have been associated with cancer (Martens-Uzunova et al., 2011; Patterson et al., 2017; Chow and Chen, 2018).

1.4.2 Transfer RNA derived small RNAs

Transfer RNAs (tRNAs) are transcribed by RNA polymerase III and are characterized by a clover leaf tertiary structure, five structural features (acceptor stem, D-loop, TΨC loop, anticodon loop, variable loop) and 5' leader and 3' trailer sequences (Kumar et al., 2016; Shen et al., 2018). During tRNA maturation the 5' leader sequences of the pre-tRNA is removed by the endoribonuclease RNase P (Frank and Pace, 1998) whereas the 3' trailer sequence is cleaved by RNase Z (ELAC2) (Schiffer et al., 2002).

This is followed by the addition of the nucleotides CCA to the processed 3' end via the tRNA nucleotidyl transferase and nuclear export to the cytoplasm (Maraia and Lamichhane, 2011).

Advances in next generation sequencing technologies and bioinformatics analysis revealed that tRNAs are cleaved into smaller RNA fragments (Lee et al., 2009). These tRNA derived sRNA fragments can be classified depending on their origin and biogenesis into stress-induced tRNA fragments termed tiRNA or tRNA halves and regulatory tRNA fragments (tRFs).

The tRNA halves are fragmented upon various stress conditions, such as oxidative stress, nutrient starvation, hypothermia, UV irradiation (Fu et al., 2009; Yamasaki et al., 2009) but are also present under normal conditions in low abundance (Thompson et al., 2008). These tRNA fragments have been initially identified in *Escherichia coli* (*E. coli*) during bacteriophage infection. Later these fragments were found in *Tetrahymena thermophila* during nutrient starvation (Lee and Collins, 2005). From then on, tRNA halves were discovered in more species of bacteria (Haiser et al., 2008), fungi (Jöchl et al., 2008) and eukaryotes (Thompson et al., 2008). The tRNA halves are 31-40 nt in length and are generated primarily by specific cleavage in the anticodon loop of the mature tRNA by the ribonuclease angiogenin in humans (Thompson et al., 2008; Yamasaki et al., 2009) and Rny1p in yeast (Thompson and Parker, 2009). Depending on the position of the cleavage site at the anticodon loop tRNA halves can be classified into 3' and 5' end derived tRNA halves (Thompson et al., 2008; Kumar et al., 2016). It was shown that tRNA halves are involved in the inhibition of translation independent of the phosphorylation of the translation initiation factor eIF2α by displacing eIF4G/eIF4A from mRNA (Ivanov et al., 2011) and formation of stress granules (Emara et al., 2010). Interestingly, only the 5' end derived tRNA halves and not the 3' end derived tRNA halves were able to repress protein translation (Yamasaki et al., 2009; Emara et al., 2010).

Apart from stress-induced tRNA halves, next generation sequencing of prostate cancer cell lines revealed that there are three classes of regulatory tRFs, tRF-1, tRF-3, tRF5 that are between 17-26 nt in length (Lee et al., 2009). In addition to these three described classes of tRFs internal tRFs (itRFs) have been recently described. It was shown that itRFs are generated by a combination of cleavages in the anticodon loop and either D arm or TΨC arm of the tRNA (Keam and Hutvagner, 2015; Telonis et al., 2015).

The tRF-1 fragments are generated by removal of the 3' trailer of a pre-tRNA by cleavage mediated by the ribonuclease RNaseZ/ELAC2 (Lee et al., 2009). The tRNA fragment tRF-1001 was the most abundant tRF in the prostate cancer lines. In knockdown experiments it was observed that tFR-1001 was essential for cell proliferation and was required in the transition between G2 and M phase during cell cycle (Lee et al., 2009).

It is reported that Dicer and other ribonucleases such as angiogenin are involved in the specific cleavage and generation of tRF-3 and tRF-5 fragments. The tRF-3 fragments are either ~18 nt or ~22 nt in length and are produced by cleavages in the TΨC loop. The tRF-5 fragments vary in length between 14-30 nt and are cleaved at the D-loop or the stem region between the D-loop and anticodon loop (Kumar et al., 2014). Several studies showed that the processing of tRFs can be Dicer dependent (Cole et al., 2009; Lee et al., 2009; Haussecker et al., 2010; Pederson, 2010) or can occur Drosha and Dicer independent (Li et al., 2012; Kumar et al., 2014). However, the ribonucleases and mechanisms involved in the generation of tRFs still remain elusive (Kumar et al., 2014).

Next generation sequencing experiments of sRNAs immunoprecipitated with Ago1, Ago2, Ago3 and Ago4 in human cells revealed that tRF-3 fragments preferentially associate with Ago1, Ago3 or Ago4, instead of Ago2 (Haussecker et al., 2010; Kumar et al., 2014). It was observed that some tRFs are involved in RNA silencing mechanisms and act in a similar way to miRNAs and repress translation (Haussecker et al., 2010). However, only tRF-5 fragments have been shown to be involved in gene silencing mechanisms and translational repression (Sobala and Hutvagner, 2013).

Depending on their association with different Ago proteins tRFs function in several biological processes such as cell growth, cell proliferation, apoptosis, DNA damage response, assembly of stress granules, genome stability, epigenetic inheritance and various diseases such as cancer (Shen et al., 2018; Zhu et al., 2018).

Recently it was shown that the endoribonuclease RNase L is responsible for tRNA cleavage in response to viral infection in human A549 lung cancer cells. Donovan *et al*. demonstrated that dsRNA induced tRNA cleavage by RNase L resulted in translation arrest and inhibition of protein synthesis (Donovan et al., 2017).

1.4.3 Vault RNA derived small RNAs

Bioinformatics analysis of next generation sequencing experiments led to the discovery of vault RNA derived sRNAs (svRNAs). Vault RNAs have a characteristic stem loop structure and produce fragments from the 3' end and 5' end of the terminal stem. Persson *et al*. demonstrated that human svRNAs are processed in a different way than canonical miRNAs. The generation of svRNAs was reported to be independent of Drosha, but Dicer dependent. Interestingly, one svRNA was reported to be loaded into Ago2 and act like a miRNA by regulating the gene expression of CYP3A4 which is involved in drug resistance (Persson et al., 2009).

With the recent advances in next generation sequencing technologies vault RNA derived sRNAs have been found highly enriched in extracellular vesicles derived from immune cells (Nolte-'t Hoen et al., 2012).

Recently, it was shown that RNA modifications play an important role in the processing of vault RNAs into smaller RNA fragments. It was reported that the RNA methyltransferase NSun2 is responsible for the cytosine-5 methylation (m^5C) of vault RNAs which affects the generation of svRNAs. In this study it was demonstrated that svRNA4 associates with Ago2 and Ago3 and acts in a miRNA like manner (Hussain et al., 2013). In humans the loss of NSun2 is correlated with neurodevelopmental diseases (Abbasi-Moheb et al., 2012; Martinez et al., 2012). Interestingly, in NSun2 deficient cells in which the processing into svRNAs is affected an increase of mRNAs related to neurodevelopmental disorders was observed. This suggested that the altered processing of vault RNAs into svRNAs in NSun2 deficient cells may contribute to developmental disorders (Hussain et al., 2013).

1.5 Y RNAs

1.5.1 Y RNAs in general

Another class of non-coding RNAs that generate sRNA fragments are Y RNAs which are subject of this thesis. Y RNAs were initially discovered in 1981 as components of the Ro60 and La ribonucleoprotein complexes (RoRNP) in sera of patients with the autoimmune disorders systemic lupus erythematosus (SLE) and Sjögren's syndrome (Lerner et al., 1981a; Wolin and Steitz, 1984). Y RNAs were shown to associate with the 60 kDa protein Ro60 and the 47 kDa protein La. Both Ro60 and La auto-antigens have a RNA binding motif that consists of 80 amino acids. Ro60 is conserved among vertebrates and orthologues have also been described in prokaryotes. Ro60 is mainly involved in RNA quality control, RNA stability and cellular stress response.

The La auto-antigen was identified in a subset of RoRNPs (Hendrick et al., 1981). The La protein associates with RNA polymerase III transcripts and is involved in the accurate and efficient termination of RNA polymerase III transcription. The La protein binds to the 3' polyuridine tail of newly synthesized transcripts in the nucleus.

Y RNAs were discovered by Lerner and colleagues who identified small cytoplasmic RNAs by immunoprecipitation of total cell extracts from mouse Ehrlich ascites and serum from SLE patients containing antibodies against Ro and La proteins. The prefix "Y" was attributed to this class of non-coding RNAs and these RNAs were termed Y RNAs because of their cytoplasmic localization in contrast to snRNAs (URNAs) which were found in nuclear extracts of Ehrlich ascites cells. Hendrick *et al*. confirmed by immunofluorescence experiments that Y RNAs are cytoplasmic (Hendrick et al., 1981).

Y RNAs are transcribed by RNA polymerase III from single copy genes and range in length between 84-112 nt (Hendrick et al., 1981; Wolin and Steitz, 1983; Maraia et al., 1994a). Human Y RNAs were found highly abundant in heart and brain tissues and low abundant in liver tissues (Wolin and Steitz, 1984).

1.5.2 Evolution of Y RNAs

Y RNAs have been identified so far in vertebrates, insects, nematodes and their orthologues have been described in prokaryotes, but not in yeast or higher plants. In several green algae such as *Chlamydomonas reinhardtii*, *Volvox carteri* and *Tetrabena socialis* as well as at least one fungus *Allomyces macrogynus* Ro60 orthologues have been identified (Wolin, 2019).

Y RNAs are highly evolutionarily conserved among vertebrates and the number of distinct Y RNAs varies between one and four Y RNAs. Y3 seems to be the most conserved Y RNA among all Y RNAs (Farris et al., 1996). It was proposed that all Y RNA genes are derived from a common single ancestral Y RNA gene (Wolin and Steitz, 1983). It was suggested that gene duplication events resulted in the development of Y1/Y3 and Y4/Y5 paralogues (Mosig et al., 2007).

In humans there are four Y RNAs, named hY1, hY3, hY4 and hY5 RNA, while hY2 is a truncated version of hY1 (Wolin and Steitz, 1983). In contrast to humans, rodents only possess Y1 and Y3 homologues (mY1 and mY3), while they have a redundant “fossil” Y5RNA gene which is not expressed anymore (Pruijn et al., 1993).

Moreover, bioinformatics analysis revealed that humans and chimpanzees contain more than 1000 Y RNA pseudogenes, whereas mice only have a few pseudogenes (Perreault et al., 2005; Perreault et al., 2007). Sequence analysis of the flanking regions suggested that the human Y RNA pseudogenes were generated by the integration of mutated Y RNAs through L1 retro transposition. Interestingly, hY1 and hY3 gave rise to the majority of hY RNA pseudogenes whereas hY5 RNA resulted only in a few pseudogenes (Perreault et al., 2005).

Sequence homology analysis of Y RNAs and Y RNA pseudogenes across various different metazoans suggested that the different Y RNA species evolved from a common single ancient ancestor Y RNA gene through tandem gene duplication events. This original ancient Y RNA gene might have diverged into two putative Y RNAs in nematodes and Y1 and Y3 in the ancestor of fish. The Y4 and Y5 RNA gene might have evolved later in evolution and were first present in amphibians. Thus, there exist four major clades of Y RNAs Y1, Y3, Y4 and Y5 in vertebrates, including the teleost Y1/Y3 RNAs and amphibian Y5 RNAs (Mosig et al., 2007; Perreault et al., 2007; Duarte Junior et al., 2019).

In humans, all Y RNA genes are located and clustered together within 45 kb at a single locus on chromosome 7q36 and all Y RNA genes are transcribed by RNA polymerase III from distinct promoters.

Y3, Y4 and Y5 RNA are transcribed from the plus strand whereas Y1 is transcribed from the minus strand (Maraia et al., 1994b; Maraia et al., 1996, Hendrick et al., 1981; Wolin and Steitz, 1983).

The four Y RNA genes are highly conserved among vertebrates and the number of Y RNA genes varies among species (Pruijn et al., 1993). Frogs have four Y RNA genes xY3, xY4, xY5 and Yα (O'Brien et al., 1993). Interestingly the Y1 gene might have gone lost during evolution and through gene duplication of the Y5 gene the paralogue Yα gene was gained (Perreault et al., 2007).

In *C. elegans* one ceY RNA was identified (Van Horn et al., 1995). However, it was shown that this ceY RNA is not essential as nematodes are viable when the ceY RNA gene was deleted (Boria et al., 2010). In addition, the ceY RNA is not involved in the initiation of DNA replication and has no structure and/or sequence similarity to vertebrate Y RNA (Boria et al., 2010; Gardiner et al., 2009). Thus, ceY RNAs might be involved in other cellular processes in nematodes.

Interestingly, a class of small non-coding stem bulge RNAs (sbRNAs) has been discovered in nematodes (Deng et al., 2006; Aftab et al., 2008; Boria et al., 2010).

During recent years sbRNA homologues have been identified in the insect *Anopheles gambiae* (Perreault et al., 2007), the silkworm *Bombyx mori* (Duarte Junior et al., 2015) or *Branchiostoma floridae* (Mosig et al., 2007).

Interestingly, these sbRNAs were identified as homologues of vertebrate Y RNAs and are conserved in both their sequence and secondary structure. Phylogenetic analysis showed that nematode sbRNAs are not related to Y1, Y3 or Y4, but are more closely related to Y5. Intriguingly, the sbRNAs identified in *Bombyx mori* contained a highly evolutionarily conserved sequence motif (GTGGCTTATC) with hY5 RNA and Chinese hamster chY5 RNA. This suggested that these sbRNAs may function as Y RNAs but further investigations are needed in the functions of sbRNAs in *Bombyx mori* (Duarte Junior et al., 2015; 2019).

Y RNAs and sbRNAs are expressed in vertebrates and nematodes around ten times less than ribosomal 5S RNAs (Christov et al., 2006). It was shown that most sbRNAs were differentially expressed during early embryonic development. Interestingly, in

adult worms, dauer larvae and nematodes under heat stress most of the sbRNAs were upregulated (Deng et al., 2006, Boria et al., 2010).

Recently Kowalski *et al.* reported that sbRNAs are essential for cell proliferation and early embryonic development of *C. elegans*. Knockout of sbRNAs by antisense morpholino oligonucleotides resulted in defects during the initiation of DNA replication and increased lethality of *C. elegans* embryos (Kowalski et al., 2015).

Both nematode sbRNAs and Y RNAs have a characteristic stem-loop structure with a bulged double-stranded stem domain and a pyrimidine-rich single-stranded loop domain. The single-bulged cytosine separates the lower stem from the upper stem domain. The upper stem domain is characterized by a highly conserved A/GUG-CAC-U nucleotide sequence motif which was shown to be essential for the initiation of chromosomal DNA replication (Boria et al., 2010; Christov et al., 2006; Gardiner et al., 2009; Kowalski et al., 2015; Wang et al., 2014). Interestingly, both sbRNAs and Y RNA have a central loop domain that allows flexibility. Remarkably, this central loop also includes a highly conserved sequence motif of UUAUC at the 5' end of the loop domain downstream of the GUG-motif. By substitution mutations of this penta-nucleotide motif UUAUC it was shown that the upper stem domain and the UUAUC motif downstream of the GUG sequence motif of sbRNAs are important for the initiation of DNA replication (Boria et al., 2010; Wang et al., 2014; Kowalski et al., 2015).

Recently, two functional sbRNAs named Dm1 and Dm2 were identified and characterized in *Drosophila melanogaster* for the first time. Phylogenetic analysis showed that these sbRNAs are evolutionarily closer related to nematode sbRNAs than to vertebrate Y RNAs. Interestingly, Dm1 and Dm2 sbRNAs have similar secondary structure features to nematode sbRNAs and both contain a GUG-CAC motif and a UUUAC penta-nucleotide motif which is similar to the UUAUC motif in vertebrate Y RNA sequences. Intriguingly, it was shown that Dm1 and Dm1 sbRNA, but not Dm2 sbRNA was able to initiate DNA replication in a human cell-free system when vertebrate Y RNA were depleted. Thus, it was suggested that Dm1 sbRNA has similar functions than vertebrate Y RNAs and can be considered as a functional homologue of vertebrate Y RNAs (Duarte Junior et al., 2019).

Y RNA have also been discovered in various prokaryotes such as *Deinococcus radiodurans* (*D. radiodurans*) (Chen et al., 2000), *Salmonella enterica serovar typhimurium (S. typhimurium)* (Armour et al., 2009) or *Mycobacterium smegmatis* (Chen et al., 2014).

1.5.3 Secondary structure of Y RNAs

All human Y RNAs show evolutionary conservation in their sequence and secondary structures and fold into characteristic stem-loop structures (Wolin and Steitz, 1983). The secondary structures of human Y RNAs were initially predicted based on sequence alignment and computational minimum-free energy (Kato et al., 1982; Wolin and Steitz, 1983, 1984).

However, these minimum-free energy based calculated predictions were limited and some important structural features were further determined using both enzymatic cleavage and nucleotide chemical modification experiments (Wolin & Steitz, 1984; van Gelder *et al.*, 1994).

Van Gelder *et al*. investigated the structural features of hY1 and hY5 biochemically in more detail by further enzymatic cleavage and chemical probing of hY3 and hY4 RNA. The human Y RNAs were digested with several RNases under native and denaturing conditions in order to determine stacked, double-stranded or single-stranded RNA regions. The RNase V1 cleaves double stranded RNAs (dsN), RNase A cleaves single-stranded RNA regions after C and U (ssC and ssU), RNase T1 cleaves single-stranded RNAs after G (ssG) and RNase T2 single-stranded RNAs (ssN) (van Gelder et al., 1994).

In the chemical probing experiments the nucleotides were chemically modified by dimethylsulfide (DMS), 1-cyclohexyl-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate (CMCT) and kethoxal treatment. DMS mainly reacts with the N1 atom of adenosines (N1-A), while CMCT modifies N3-U and kethoxal is reactive with N1-G and N2-G. Primer extension analysis was performed to identify the chemically modified nucleotides and get the exact positions of all Watson-Crick bases (van Gelder et al., 1994).

The structure probing experiments of hY1 and hY5 revealed that the secondary structures of hY1 and hY5 were in general similar to the minimum-free energy calculated secondary structure predictions of hY1 and hY5 RNA (van Gelder et al., 1994). However, the pyrimidine rich region of hY1 RNA was not found single stranded as previously anticipated from earlier computational based sequence alignment (**Figure 1.1**). The hY1 RNA was resistant to structure probing chemicals and enzymes under native conditions but not under semi-denaturing conditions. The authors suggested that tertiary interactions or base stacking were involved which might be important for RNA binding protein recognition (van Gelder et al., 1994).

Intriguingly, under native conditions the hY5 RNA formed a large internal loop which was contradictory to the minimum-free energy based calculated secondary structure (shown in **Figure 1.1 A**). In this computational prediction using Mfold (**Figure 1.1 A**) it was suggested that base pairing between U23-G46 (**Figure 1.1**, shown in red) as well as between A24-U45 (**Figure 1.1**, shown in green) occurred. This base pairing resulted in the formation of a smaller internal loop and a mismatch between U25-U44 (Pruijn et al., 1991) whereas in the enzymatic and chemical structure probing experiments there was no base pairing between U23 and G46 (**Figure 1.1 B**, shown in red) and A24 and U45 (**Figure 1.1 B**, shown in green) and an internal loop was formed (van Gelder et al., 1994).

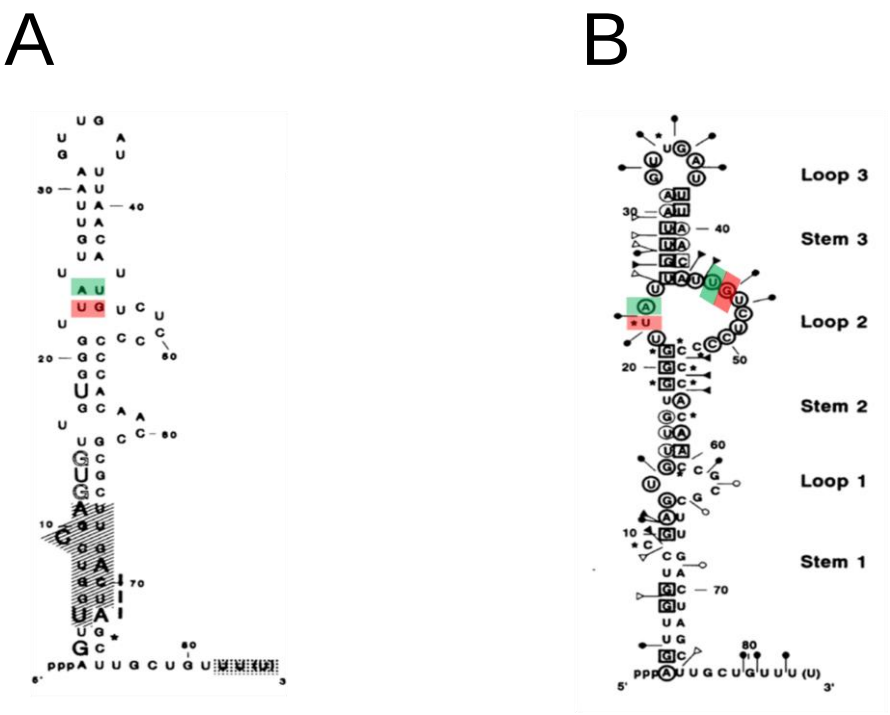


Figure 1.1. Secondary structure predictions of hY5 RNA using Mfold and chemical structure probing experiments.

A, Predicted secondary structure of hY5 RNA using Mfold after Pruijn *et al*., 1991. The predicted structure using Mfold suggested that base pairing between U23 and G46 (shown in red) and A24 and U45 (shown in green) occurred, which resulted in a smaller internal loop and a mismatch between U25 and U44. The shaded region shows the highly conserved Ro60 binding site whereas the dotted region indicates the La binding site. The figure was taken from Pruijn *et al.*, 1991 and modified.

B, Enzymatic and chemical structure probing of hY5 RNA after van Gelder *et al*., 1994. In contrast to the computationally predicted secondary structure of hY5 using Mfold (A) the enzymatic and chemical structure probing experiments indicated that under native conditions the internal loop 2 is completely accessible and no base pairing between U23 and G46 (shown in red) and A24 and U45 (shown in green) was observed.

Bold and light circles around the individual nucleotides mean strong and moderate reactivity to DMS under native conditions. Squares around the nucleotides indicate the nucleotides that were unreactive under native conditions, but showed reactivity towards CMCT under semi-denaturing conditions. Asterisks mark the positions where the reverse transcription was stopped. The enzymatic cleavage sites of single strand-specific RNases (T1, A, T2) under native conditions are marked by small solid circles and are bold for strong reactivity and open circles for moderate reactivity. The cleavages by RNase VI are indicated by triangles and a bold triangle symbol means strong reactivity whereas open triangle suggests moderate reactivity. The figure was taken from van Gelder *et al*., 1994 and modified.

In a later study performed by Teunissen *et al*. the secondary structures of Y3 and Y4 RNA from human, frog and iguana were investigated *in vitro* using enzymatic and chemical probing experiments and compared with phylogenetic data. The obtained structure probing data for Y3 and Y4 RNA were partly different compared to the computationally predicted secondary structures but all Y RNAs showed highly conserved structural elements in human, frog and iguana Y RNAs (Teunissen et al., 2000).

Previous phylogenetic analysis using a novel secondary structure prediction algorithm also revealed that the secondary structure of Y1, Y3 and Y4 RNA was highly conserved in rabbit, duck, trout, guinea pig and cow suggesting that the secondary structure of Y RNAs is conserved among vertebrates (Farris et al., 1999).

All Y RNAs are characterized by at least two conserved stem structures. The stem formed by extensive base pairing between the 3' and 5' region of the Y RNAs is the most conserved. This region contains a single bulged cytosine within a conserved helix and an internal loop which separates the lower stem from the upper stem (**Figure 1.2**, **A**).

Y RNAs are characterized by an internal single-stranded pyrimidine-rich loop domain which is least conserved and varies in nucleotide sequence and size between 13-36 nt. From the structure probing experiments it was suggested that the central part of Y RNAs is rather dynamic and accessible to interaction partners such as other nucleic acids or proteins. This would suggest that this region might be important for the functionality of Y RNAs (van Gelder et al., 1994; Teunissen et al., 2000). In overall, eukaryotic and prokaryotic Y RNAs are modular and the binding site of Ro60 orthologues in the lower stem region is highly conserved among vertebrates, nematodes and prokaryotes **(Figure 1.2, B).**

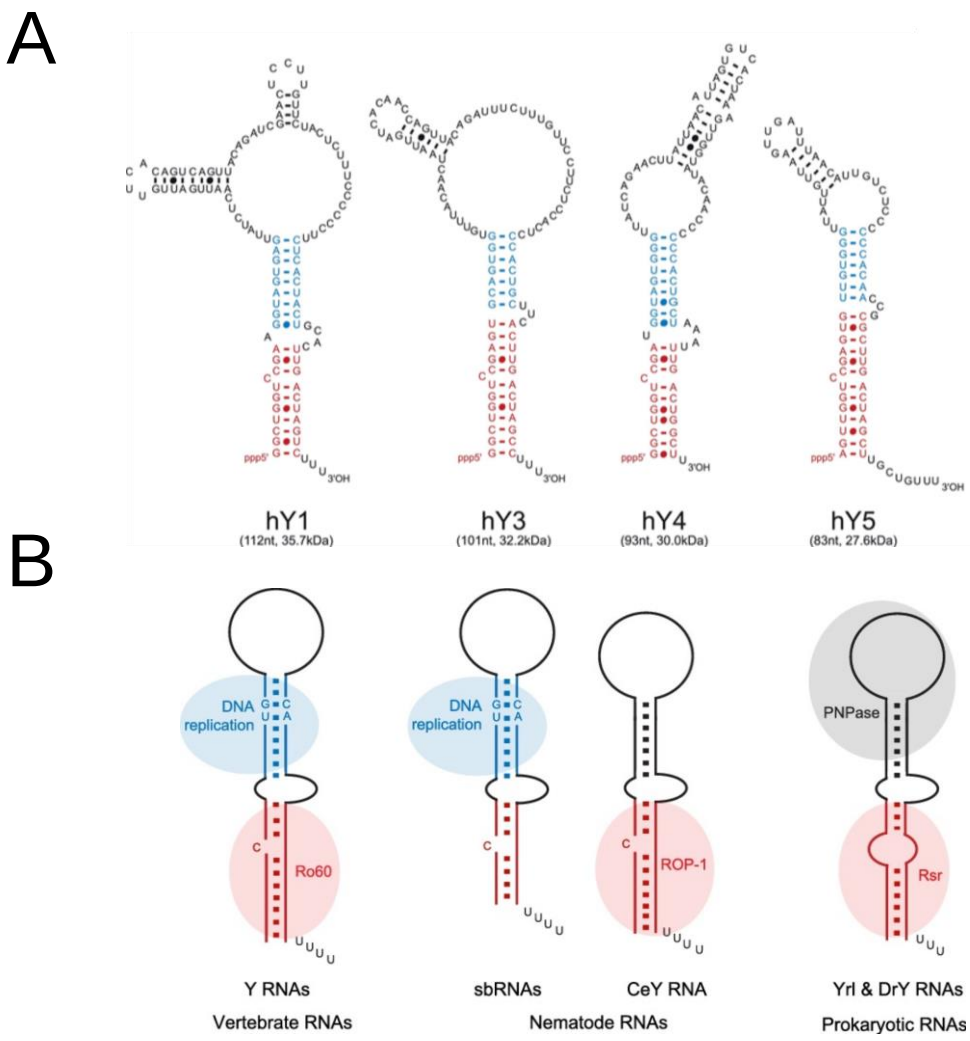


Figure 1.2. Secondary structures and modular composition of eukaryotic and prokaryotic Y RNAs.

A, Secondary structures of human Y RNAs. Human Y RNAs are evolutionarily conserved in their sequence and secondary structure. The region highlighted in red shows the conserved stem domain where Ro60 binds. The blue highlighted upper stem was shown to be involved in DNA replication.

B, Modular structure of Y RNAs. The region shown in red is the highly conserved binding site of Ro60 orthologues of vertebrate, nematode and prokaryotic RNAs that is composed of a lower stem that is formed by extensive base pairing of the 3' and 5' end. The upper stem is highlighted in blue and plays an essential role in DNA replication. The internal single-stranded pyrimidine-rich loop domain is least conserved. In prokaryotes the exoribonuclease polynucleotide phosphorylase PNPase shown in grey interacts with the Ro60 orthologue Rsr forming the RNA degradation complex RYPER (Ro60/Y RNA/PNPase exoribonuclease RNP). The figure was taken from Kowalski and Krude 2015.

Ribonuclease protection experiments revealed that Ro60 binds to the lower stem region of the conserved stem that contains a bulged helix. Deletion experiments showed that a single cytidine bulge is essential for Ro60 binding (Wolin and Steitz, 1984; Pruijn et al., 1991; Green et al., 1998) (**Figure 1.2**, highlighted in red).

Green *et al.* demonstrated by mutagenesis and chemical structure probing experiments that Ro60 recognizes specific base pairs within the bulged conserved helix and binds to the major groove of Y RNAs. It was observed that a single bulged nucleotide and a bulge of 3 nt on the opposite strand of the helix are required for Ro60 binding. In addition, it was shown that the base pairing between C9 and G90 is critical for Ro60 binding and the bulge at C8 is required for Ro60 recognition (Green et al., 1998).

In previous structure probing experiments of naked hY1 and hY5 RNAs performed by van Gelder *et al*. it was reported that in the majority of Y RNA molecules the cytosine at position 9 was bulged out.

However, in a few molecules base-pairing between C9-G102 occurred and C8 was bulged out which suggested that two conformations of Y RNAs bound to Ro60 exist (**Figure 1.1 B**) (van Gelder et al., 1994).

Interestingly, Green *et al*. performed structural analysis of naked and protein-bound RNA and found that Ro60 binding resulted in a structural change of Y RNAs. From these results it was suggested that in the naked Y RNA the cytidine at position 9 is unpaired whereas in the Ro60 bound form the cytidine at position 8 is bulged out and a base pair between C9 and G90 in the xY3 RNA is formed (Green et al., 1998).

Furthermore, Y RNAs seem not have posttranscriptional modifications (Kato et al., 1982) and are characterized by an oligo-uridine stretch at the 3’ terminus which is the binding site for the La protein. At the 5’ end Y RNAs have a tri-phosphorylated guanine residue (pppG), whereas hY5 shows a tri-phosphorylated adenine residue (pppA) (Hendrick et al., 1981; Kowalski and Krude, 2015).

1.6 Y RNAs in RoRNPs

1.6.1 Ro60 protein

During investigations of the autoimmune disorders SLE and Sjögren's syndrome it was shown that the 60 kDa protein Ro60 (SSA) and the 47 kDa protein La (SSB) are the major autoantigens that associate with Y RNAs forming RoRNPs. It was found that in patients with SLE, anti-Ro antibodies are linked to photosensitivity and neonatal lupus in which mothers give birth to children suffering from cardiac conduction defects or complete heart block. Xue *et al*. showed that mice lacking Ro60 developed a lupus-like syndrome which proposed that Ro60 might play a role in the prevention of autoimmunity (Xue et al., 2003).

Ro60 is evolutionarily conserved among vertebrates and has been characterized in mammalian cells (Deutscher et al., 1988). Its orthologues have been described in *Xenopus laevis* (O'Brien et al., 1993) and nematodes (Van Horn et al., 1995; Labbé et al., 1999) and prokaryotes such *D. radiodurans* and *Salmonella* (Wolin et al., 2013). Although RoRNPs have been detected in sera of SLE patients over 30 years ago, the precise functional role of RoRNPs and Y RNAs has yet to be elucidated.

The crystal structure of full-length *Xenopus laevis* Ro60 that was resolved with and without ligands revealed that Ro60 has a doughnut-shaped structure. Ro60 has two domains. One domain of the Ro60 protein shows a similar structure to the van Willebrand factor A (vWFA) domain whereas the second domain is composed of several α-helical HEAT repeats that form the toroid structure and a central cavity which is positively charged. It was shown that the single stranded 3' ends of misfolded RNAs bind via this central cavity and not dsRNAs. Interestingly, Y RNAs bind to the outer ring domain along with misfolded RNAs. Y RNAs might sterically prevent binding of misfolded RNAs on the outer surface of Ro60 via competition for the same binding site. Hence, it was suggested that Y RNAs regulate the access of misfolded RNAs and other RNA molecules to the Ro60 protein (Stein et al., 2005; Fuchs et al., 2006). Further, a metal ion dependent adhesion site (MIDAS) motif was identified in the vWFA domain. This motif was proposed to be responsible for the conformational change of Ro60 during opening and closing of the toroid structure. Therefore, this motif might also be involved in the recruitment of other protein interaction partners or ribonucleases and regulate the accessibly of misfolded RNAs to Ro60 (Stein et al., 2005).

1.6.2 Role of RoRNPs in RNA quality control

Several studies revealed that Ro60 is involved in quality control pathways of misfolded RNAs. Immunoprecipitation experiments of *X. laevis* oocyte nuclei showed that Ro60 associates with defective 5S rRNA precursors (O'Brien and Wolin, 1994). These variants of 5S rRNA contain an extra 8-10 nt at the 3' end which are generated by a failure of RNA polymerase III transcription termination. The misfolded 5S rRNA precursors cannot be efficiently processed to mature 5S rRNAs. The variant 5S rRNA molecules fold into alternative secondary structures, which are recognized by Ro60 resulting in the degradation of defective 5S rRNA precursors (O'Brien and Wolin, 1994; Shi et al., 1996).

In mouse embryonic stem cells it was observed that Ro60 also binds misfolded U2 small nuclear RNAs (snRNAs) that have important functional roles in pre-mRNA splicing. When wild type and variant U2 small nuclear RNAs were injected into *X. laevis* oocytes at the same time, Ro60 associated with the misfolded U2 snRNA in the nucleus (Chen et al., 2003).

The binding of Ro60 to misfolded 5S rRNA precursors in *X. laevis* oocyte nuclei and variant U2 snRNAs in mouse embryonic stem cells in *X. laevis* oocytes suggests that Ro60 plays an important role in quality control pathways of small non-coding RNAs.
One Y RNA homologue, called ceY RNA was identified in *C. elegans* which associates with ROP-1. Intriguingly, ceY RNAs were drastically downregulated in *rop-1* mutants which indicates that ceY RNAs are stabilized by binding to the Ro protein (Labbé et al., 1999). Also, in mouse cells it was shown that the number of Y RNAs was decreased in cells lacking Ro60 which strengthens the possibility that binding to Ro stabilizes Y RNAs (Chen et al., 2000).
Interestingly, Hogg and Collins suggested that Y RNAs recruit defective and misfolded RNAs to Ro60. They showed that human Y5 RNAs interact with variant 5S rRNAs via the ribosomal protein L5. It was observed that Y5 RNA has a preference for binding to 5S rRNA variants compared to the correctly processed 5S rRNA transcripts. Consequently, the Y5 RNA might assist Ro60 to differentiate between correctly folded and misfolded 5S rRNAs, suggesting that it facilitates the access of RNA variants to the outer surface of Ro60 (Hogg and Collins, 2007).

1.6.3 Role of RoRNPs in stress response

Ro60 is also involved in the survival of cells upon various types of environmental stress, such as UV irradiation in bacteria and animal cells (Sim and Wolin, 2011; Chen et al., 2014). An orthologue of Ro60, called Rsr (Ro-sixty related) and a single Y RNA have been identified in the radiation-resistant bacteria *D. radiodurans.* When the *D. radiodurans* strains were subjected to UV irradiation both the Ro60 homologue Rsr and the associated Y RNA were upregulated in the cells. Bacterial strains lacking Rsr (Δrsr) were shown to be more sensitive following UV irradiation, compared to the wild type *D. radiodurans* strains (Chen et al., 2000). Both Rsr and Y RNA were accumulated upon different environmental stress conditions such as ionizing radiation, heat stress or during the stationary phase of *D. radiodurans* growth. In addition, mouse cells lacking the Ro protein were more sensitive to UV irradiation compared to normal cells and both Ro60 and mY3 were found accumulated in the nucleus (Chen et al., 2003).

A further important finding in *D. radiodurans* revealed that during stress Y RNAs and Ro are involved in maturation and decay pathways of structured RNAs including exoribonucleases (Wurtmann and Wolin, 2010). It was shown that Y RNAs tether the Ro60 ortholog Rsr to the 3' to 5' exoribonuclease polynucleotide phosphorylase PNPase forming the RNA degradation machinery RYPER (Ro60/Y RNA/PNPase Exoribonuclease RNP). The ring shaped Rsr interacts with single-stranded structured non-coding RNAs via its central cavity where PNPase degrades structured RNAs. Rsr interacts with the exoribonucleases RNase II and RNase PH during heat stress in 23S rRNA maturation. Biochemical experiments showed that Y RNA binding to Rsr inhibits rRNA maturation. Ribosomal RNAs are only processed when Y RNAs are not bound to Rsr. Y RNAs tether the Ro60 ortholog Rsr to other proteins in order to form either the degradation complex RYPER in rRNA decay pathways or to interact with other exoribonucleases in rRNA maturation during stress (Chen et al., 2007; Chen et al., 2013; Wolin et al., 2013; Chen et al., 2014).

In conclusion, RoRNPs are involved in quality control pathways and RNA stability. As part of RoRNPs Y RNAs play an important role in the regulation of the access of other RNAs to the central cavity of Ro and can be seen as "gate keepers" (Chen et al., 2013). Both Ro60 and Y RNAs are important for cell survival under stress conditions (Chen et al., 2000). However, the function of Ro60 and Y RNAs correlated with the pathogenesis of autoimmune disorders is still not clear.

1.6.4 Subcellular localisation of RoRNPs

Intriguingly, the Ro60 protein has been localized in both the cytoplasm and nucleus. In contrast, Y RNAs are primarily localized in the cytoplasm of mammalian cells and *X. laevis* oocytes (O'Brien et al., 1993; Peek et al., 1993). Simons *et al*. suggested after microinjection of *X. laevis* oocytes and immunoprecipitation experiments that Ro60 and La protein are transported into the nucleus after translation and associate with Y RNAs in the nucleus where the RoRNP assembly occurs (Simons et al., 1994). The RoRNP complex is then exported to the cytoplasm. For the nuclear export of Y RNAs into the cytoplasm the Ro60 binding site needs to be intact (Simons et al., 1996). Further it was shown that when Ro60 protein was mutated, Ro60 was primarily localized in the nucleus and not associated with Y RNAs (Sim et al., 2009). In this study it was found that Ro60 has a nuclear accumulation signal that is masked by the binding of Y RNAs to Ro60. During stress Ro60 accumulates in the nucleus and the nuclear accumulation signal becomes accessible. This suggested that Y RNAs play an important role in the regulation of the subcellular distribution of Ro60 under normal and stress conditions (Sim et al., 2009; Sim and Wolin, 2011).

A later study showed that out of all human Y RNAs, only the hY5 RoRNP was retained in the nucleus which suggested that it has an additional function that other human Y RNAs lack (Gendron et al., 2001).

Due to the different techniques that were applied the subcellular localisation of Y RNAs still remains controversial. *In situ* hybridization and electron microscopy of cultured human cells suggested that Y RNAs are present in both the nucleus and cytoplasm (Farris et al., 1997). In proliferating cells it was observed that hY1, hY3 and hY5 are localized in the perinucleolar compartment (Matera et al., 1995).

In a recent study it was reported that during DNA replication all Y RNAs interact dynamically with chromatin and replication proteins in G1 phase nuclei. This was shown *in vitro* using fluorescently labelled Y RNAs in a human cell-free system. Interestingly, hY1, hY3 and hY4 were associated with early replicating euchromatin, whereas hY5 RNA was primarily found in nucleoli (Zhang et al., 2011).

1.7 Y RNA binding proteins

Using biochemical approaches several stable or transient Y RNA associated proteins have been identified. Interestingly, most of the described Y RNA binding proteins have been shown to be involved in alternative splicing and translation.

1.7.1 La protein

Apart from Ro60 the La protein (SSB) is a major autoantigen which was found in sera of patients with SLE and Sjögren's syndrome in complexes with Y RNAs and a subset of RoRNPs (Alspaugh and Tan, 1975; Hendrick et al., 1981). The La protein is a phosphoprotein that is highly abundant and primarily found in the nucleus under normal conditions, but is also present in the cytoplasm (Pruijn et al., 1991; Slobbe et al., 1992). Compared to Ro60 the La protein is almost 50 times more abundant (Peek et al., 1993). The human La protein is 47 kDa in size, comprised of 408 amino acids and highly evolutionarily conserved (Wolin and Cedervall, 2002).

The La protein binds all RNA polymerase III transcripts such as pre-tRNAs, pre-5SRNAs, U6 snRNAs, 7SL RNAs, Alu elements, Y RNAs (Hendrick et al., 1981). Some viral encoded RNAs including the Epstein-Barr virus encoded RNAs EBER 1 or EBER 2 or the adenovirus- RNAs VAI and VAII were also shown in association with La protein (Lerner et al., 1981b; Rosa et al., 1981; Rinke and Steitz, 1982; Wolin and Cedervall, 2002). The La protein is involved in the accurate termination of RNA polymerase III transcription. It was shown that La binds to the 3' polyuridine tail in the nucleus and stabilizes newly synthesized RNA polymerase III transcripts (Stefano, 1984; Gottlieb and Steitz, 1989; Maraia et al., 1994a).

Interestingly, it was shown that La is important for the nuclear retention of Y RNAs in order to protect them from degradation by exonucleases (Simons et al., 1996). La binds to the 3' polyuridine tail and increases Y RNA stability (Gottlieb and Steitz, 1989). Notably, after transcription termination Y RNAs stay associated with the La protein whereas other RNA polymerase III transcripts dissociate from it (Boire and Craft, 1989; Peek et al., 1993) suggesting that the La protein might be important for Y RNA function or recruitment of other proteins.

It was observed that during apoptosis the La protein is rapidly dephosphorylated and 25% of La proteins are cleaved at the C-terminus in a caspase dependent manner. The La protein is a phosphoprotein and synthesis of new RNA polymerase III transcripts is regulated by a phosphorylation and dephosphorylation cycle.

The induction of apoptosis results in a dysregulation of this phosphorylation cycle. This suggests that dephosphorylation of La might be important in the initiation of cell death mechanisms (Rutjes et al., 1999b).

Interestingly, the La protein is involved in the inhibition of interferon induced antiviral response. through the protein kinase PKR that gets activated by double stranded RNA. The La protein is involved in the unwinding of dsRNA which prevents the activation of the protein kinase PKR through dsRNA. (Xiao et al., 1994). Therefore, it was proposed that the dephosphorylation and cleavage of the La protein during apoptosis might be a key step to activate PKR and interferon induced antiviral response mechanisms (Rutjes et al., 1999b).

1.7.2 Nucleolin

Nucleolin is a phosphoprotein localized in the nucleolus and plays key roles in chromatin structure, rRNA biogenesis and maturation, ribosome assembly and nucleocytoplasmic transport. Nucleolin was shown to be associated with cell proliferation and cell growth (Ginisty et al., 1999; Srivastava and Pollard, 1999; Fouraux et al., 2002).

In order to identify novel RoRNP associated proteins Ro60 and La immunoprecipitation experiments and mass spectrometry were performed. In this *in vitro* experiment Nucleolin was identified as a new RoRNP associated protein. Surprisingly, nucleolin was only in complex with hY1 and hY3, but not with hY4 and hY5.

Deletion mutants on hY1 RNA revealed that a pyrimidine-rich internal loop structural feature and not the nucleotide sequence is required for efficient nucleolin association. The hY3 RNA has a similar pyrimidine-rich internal loop to hY1 RNA. However, hY4 and hY5 lack this structural feature and therefore cannot bind nucleolin.

It was hypothesized that nucleolin is involved in the shuffling of hY1 and hY3 from the nucleus to the cytoplasm since nucleolin was reported to be involved in nucleocytoplasmic transport (Fouraux et al., 2002).

Controversially, Langley *et al*. observed using a human cell-free system and immunoprecipitation experiments that nucleolin binds all four human Y RNAs but preferentially associates with hY1. This result might be due to the fact that different experimental set ups (Northern blotting versus quantitative reverse transcription-PCR) and antibodies were used for this study.

Further, it was shown that Ro60, La and nucleolin are not involved in the initiation of DNA replication. Intriguingly, using real-time PCR it was observed in the same study that 50% of all Y RNAs were not associated with Ro60, La or nucleolin (Langley et al., 2010). Langley *et al*. suggested that naked Y RNAs that are not bound to proteins are involved in DNA replication *in vitro*, whereas the protein bound Y RNAs fulfil different functions including RNA quality control, RNA stability or stress response (O'Brien and Wolin, 1994; Chen et al., 2003; Sim et al., 2009).

Interestingly, by Garcia *et al*. it was reported that the Moloney murine leukemia virus (MLV) selectively incorporates host mY1 and mY3 RNAs without the requirement of Ro60 and La binding. This study further confirmed that Y RNAs can be functional independent of Ro60 or La (Garcia et al., 2009; Langley et al., 2010).

1.7.3 Ro52

Ro52 has been identified in sera of patients with SLE and Sjögren's syndrome. This protein was previously reported to interact with Y RNAs and to be part of RoRNPs with Ro60 and La (Ben-Chetrit et al., 1988). By immunofluorescence microscopy it was shown that Ro52 and Ro60 co-localize in both the cytoplasm and nucleus (Slobbe et al., 1991). Immunoprecipitation experiments using HeLa cell extract and monospecific antisera against Ro60, La and Ro52 indicated that Ro52 can only interact with hY1 RNA in the presence of Ro60 (Slobbe et al., 1992).

In later experiments using glycerol density gradient fractionation and co-immunoprecipitation experiments with Ro60 revealed that Ro52 was not part of RoRNPs and did not co-localize with Ro60 *in vivo* (Kelekar et al., 1994; Boire et al., 1995). Intriguingly, it was shown that Ro52 does not have any RNA recognition motif which suggested that Ro52 is not a RNA binding protein (Kelekar et al., 1994). This was further confirmed by Frank *et al*. who reported that Ro52 contained zinc finger and leucine zipper motifs in the predicted amino acid sequence. It was proposed that it was a DNA binding protein that belongs to the TRIM (tripartite motif) protein family (Frank et al., 1995; Ottosson et al., 2006). Ro52 was identified as an E3 ubiquitin ligase that plays important roles in cell proliferation, cell death and autoimmunity (Espinosa et al., 2006).

1.7.4 Heterogeneous nuclear ribonucleoproteins

In order to identify novel Y RNA-associated proteins, *in vitro* biochemical purification of RoRNPs and reconstitution experiments of biotinylated Y RNAs using HeLa S100 cell extracts were performed. In these experiments several Y RNA binding proteins ranging in size between 15-80 kDa were identified. Interestingly, all of these proteins preferentially interacted with Y1 and Y3 RNA (Fabini et al., 2000). Later the heterogeneous nuclear ribonucleoproteins hnRNP I, also named polypyrimidine tract protein (PTB) and hnRNP K were identified by mass spectrometry and the interaction with human Y RNAs was validated in vitro using deletion mutants of hY1. It was clearly seen that hnRNP I/PTB and hnRNP K preferentially associate with hY1 and hY3, but not hY4 or hY5 via the pyrimidine rich central loop like nucleolin. Using a yeast three hybrid screen the interaction of hnRNP I/PTB and hnRNP K with hY1 and hY3 was confirmed *in vivo* (Fabini et al., 2001).

The proteins hnRNP I also named polypyrimidine tract binding protein (PTBP1) and hnRNP K are implicated in nucleocytoplasmic transport and have key roles in RNA processing, export, RNA stability and translation. Interestingly, La was found to be required between the interaction of hnRNP I/PTB and hnRNP K and Y RNAs. Therefore, it was hypothesized that La might be molecular chaperone for Y RNAs in order to allow the association with hnRNP I/PTB and hnRNP K and regulation of translation (Fabini et al., 2001; Belisova et al., 2005).

1.7.5 Zipcode-binding protein

Previous studies indicated that the zipcode-binding protein ZBP1 (also called IGF2BP1 and IMP1) was associated with nucleolin, Ro60 and PTB which were described as Y RNA binding proteins above (Fabini et al., 2001; Köhn et al., 2010).

Using near infrared labelled Y RNAs and electromobility shift assay ZBP1 was identified as an RNA binding protein of Y3. ZBP1 forms a trimeric complex with Y3 and La. This ternary complex formation was specific to Y3 since the complex dissociated in the presence of Y5 (Köhn et al., 2010).

Later ZBP1 was found co-purified with mouse Ro60 by tandem affinity purification. Interestingly, it was shown that ZBP1 interacts with mY1 and mY3 like nucleolin, hnRNP I/PTB or hnRNP K via the Y RNA loop domains (Sim et al., 2012; Köhn et al., 2013).

ZBP1 was reported to be involved in the subcellular localization of Ro60 and mY3 RNA. Under stress conditions such as UV irradiation it was observed that ZBP1 is bound to the mY3/Ro60 complex and accumulates in the nucleus whereas under normal conditions Ro60 and mY3 were primarily found in the cytoplasm. Also, after depletion of ZBP1 the mY3/Ro60 complex showed high accumulation in the nucleus suggesting that ZBP1 might be important for the nuclear export of RoRNPs (Sim et al., 2012). Interestingly, in the same study it was reported that apart from ZBP1, the Y-box protein YB-1 and a RNA dependent helicase MOV10 which play key roles in cellular stress response (Lu et al., 2005) associate with Ro60 and associated RNAs (Sim et al., 2012).

1.7.6 APOBEC3G

The apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3G (APOBEC3G) belongs to the family of cytidine deaminases and catalyses the deamination of cytidine to uridine. APOBEC3G is also involved in antiviral response.

Intriguingly, APOBEC3G was reported to be part of the RoRNPs and associated with all four human Y RNAs (Chiu et al., 2006; Gallois-Montbrun et al., 2008). However, the function of APOBEC3G in RoRNPs remains obscure and it was hypothesized that APOBEC3G might function in the RNA editing of Y RNAs and regulate their function (Chiu et al., 2006; Köhn et al., 2013).

1.7.7 RoRNP binding protein 1

The Ro RNP binding protein (RoBPI) is also known as poly U binding splicing factor (PUF60). RoBPI was reported to be a DNA and RNA binding protein that plays key roles in transcription and splicing (Page-McCaw et al., 1999) and can be linked with autoimmune diseases (Fiorentino et al., 2016) and cancer (Kobayashi et al., 2016).

RoBPI was identified by a modified yeast-three hybrid assay called RNA interaction trap assay (RITA) which allows the reconstitution of tripartite RNPs. Interestingly, it was observed that RoBPI primarily interacted with hY5 RNA in the nucleus, but was also reported to bind hY1 and hY3 RNA (Hogg and Collins, 2007).

Using the yeast-three hybrid assay it was found that in human cells RoBPI only associated with recombinant Ro60 and specifically hY5 RNA together but not to Ro60 or hY5 RNA alone (Bouffard et al., 2000).

Previous studies already indicated that hY5 RNA appeared to have different biochemical properties or localization compared to the other human Y RNAs. In human and mouse cells the hY5 RNA is mostly localized in the nucleus while hY1, hY3 or hY4 RNAs are cytoplasmic. Furthermore, it was suggested that the sequence of hY5 RNA itself contributes to the nuclear retention (Gendron et al., 2001). Notably, anion exchange chromatography showed that the hY5 RNA was found in a different fraction compared to hY1, hY3 or hY4 RNAs (Boire et al., 1995). Interestingly, sera from two SLE patients with autoantibodies against Ro RNPs contained antibodies specifically to hY5 RNA (Boire and Craft, 1989; Boulanger et al., 1995). Immunoprecipitation experiments after T1 RNA digestion demonstrated that a 27 nt fragment mapping to the 5' end of hY5 RNA and a 31 nt fragment mapping to the 3' end of hY5 RNA were specifically recognized by anti-hY5 antibodies (Boulanger et al., 1995).

1.7.8 Ribosomal protein L5

RNA-based affinity purification of hY1, hY3, hY4 and hY5 RNPs combined with mass spectrometry revealed that hY5 RNP specifically associated with the ribosomal protein L5 and misfolded 5S RNA. Steitz *et al*. previously reported that L5 protein is complexed with 5S rRNA and plays a role in the ribosome assembly (Steitz et al., 1988). Mutagenesis experiments revealed that certain structural features of the internal pyrimidine loop of hY5 RNA are involved in the recognition of misfolded 5S RNA. These results suggested that Y5 RNA might be involved in RNA quality mechanism and assists to distinguish between correctly folded and misfolded ncRNAs. Further it was indicated that Y5 RNA might be involved in the recruitment of other proteins such as the interferon-induced protein with tetratricopeptide repeats 5 (IFIT5) which plays a role in innate immunity (Hogg and Collins, 2007). IFIT5 was reported to bind ssRNAs that have a triphosphate group at the 5' end like the Y5 RNA (Hendrick et al., 1981; Abbas et al., 2013; Katibah et al., 2014).

1.7.9 Poly (A) specific ribonuclease

Poly(A) specific ribonuclease (PARN) is a deadenylase and mutations in PARN are linked to severe hereditary diseases such as pulmonary fibrosis or the congenital disorders which were correlated with telomere shortening (Stuart et al., 2015). Interestingly knockdown experiments revealed that PARN is required for Y RNA stability and correct 3' end processing in human cells. During 3' end processing of Y RNAs PARN and the exosome protein EXOSC10 remove the oligo (A) tails that were added by PAPD5 and stabilize Y RNAs from degradation. The 3'- to 5' exonucleases DIS3L or TOE1 recognize oligo (A) tails and are involved in RNA degradation. Interestingly, after siRNA knockdown of PARN and EXOSC10 a decrease of Y RNA levels was observed. Further knockdown experiments confirmed that PARN, EXOSC10 and PAPD5 play a key role in the 3' end processing and stability of Y RNAs. (Shukla and Parker, 2017).

1.7.10 Cleavage and polyadenylation specificity factor

The cleavage and polyadenylation specificity factor CPSF is required for the 3' end cleavage of pre-mRNAs (Chan et al., 2011). Intriguingly, by mass spectrometry of proteins that interact with biotinylated Y RNAs and copurification experiments CPSF was identified to interact with Y3/Y3** RNA (Y3** isomer of Y3). RNA sequencing analysis and knockdown experiments confirmed that among all the tested Y RNAs only the Y3 isomer Y3** was found in association with mRNA processing factors. In particular, it was confirmed that Y3** is essential for the 3' end processing of histone mRNAs (Köhn et al., 2015).

1.8 Functions of Y RNAs

1.8.1 Y RNAs in DNA replication

The first direct cellular function of Y RNAs that has been established was that Y RNAs are required for chromosomal DNA replication in vertebrates (Christov et al., 2006).

DNA replication is a vital process in order to ensure genome stability and maintenance. During cell division cycle the chromosomal DNA first gets duplicated (S phase) which is followed by a segregation into two genetically identical daughter cells (M phase). In addition to S phase and M phase the cell cycle includes two transition gap phases that allow further cell cycle regulation. The G2 phase occurs between S phase and M phase, whereas the G1 phase is between M phase and S phase (Rao and Johnson, 1970). During G1 phase the decision if the S phase and hence cell proliferation is initiated is tightly regulated. After DNA replication is initiated (S phase) several mechanisms control that all chromosomal DNA is duplicated exactly once before chromosomal segregation into two identical daughter cells in the M phase can occur (G2 phase) (Heichman and Roberts, 1994; Nurse, 1994).

Hence, a key step of the cell division cycle is the initiation of DNA replication. The process of DNA replication is described in detail in several excellent reviews and I am giving a short overview about the most important steps (Bell and Dutta, 2002; Takeda and Dutta, 2005; Pospiech et al., 2010, Fragkos et al., 2015).

In the first step of the initiation of DNA replication the coordinated assembly of the multiprotein prereplicative complexes (pre-RC) or replication licence in the early G1 phase takes place. The formation of pre-RCs is initiated by the multiprotein origin recognition complex (ORC) containing six related proteins that bind to replication origins during all steps of the cell cycle. ORC recognizes autonomously replicating sequences (ARS) that are a few hundreds of base pairs in length and are responsible for the initiation of DNA replication.

The initiation factors Cdc6 and Cdt1 bind to the ORC at the origins of replication and allow the recruitment of a group of six related minichromosome maintenance proteins (MCM2-7) to the ORC complex resulting in the formation of the pre-RC.

Pre-RCs play an important role in the initiation of DNA replication during S phase as they are involved in the unwinding of replication origins and recruitment of DNA polymerases.

During the transition between G1 phase and S phase the pre-RC needs to get activated into active replication forks. This requires another set of replication factors. The initiation of DNA replication involves the origin unwinding, stabilization of single stranded DNA and finally the recruitment of DNA polymerases. Within the pre-RC the MCM2-7 helicase complex forms a ring around the DNA and DNA gets unwound and DNA replication is initiated (see **Figure 1.3**).

Cyclin-dependent kinase CDK2 and DBF4-dependent kinase CDC7 are required for the phosphorylation und hence activation of the replication origins, the conversion of the pre-initiation complex (pre-IC) into active replication forks during S phase.

For the activation of the pre-IC complex the initiation factors Mcm10, the cell division control protein CDC45, the ATP-dependent DNA helicase Q4 (RECQL4), Treslin, the GINS complex, the DNA topoisomerase 2-binding protein (TOPBP1), DNA polymerase ε and replication protein A (RPA) are recruited in order to unwind DNA.

After the activation of MCM2-7 helicase complex several replication proteins such as replication factor C (RFC), proliferating cell nuclear antigen (PCNA) and other DNA polymerases (DNA pol α and δ) are recruited and DNA replication is initiated.

The CMG complex is formed by CDC45, the MCM2-7 helicase hexamer complex and the DNA replication complex GINS and comprises the replisome which is only formed on activated origins. After the CMG complex gets activated the DNA is unwound and DNA replication is initiated.

The check point Serin/Threonine kinases ATR and Serine protein kinase ATM that are involved in the activation of check point kinases CHK1 and CHK2 are involved in the check point control of origin activation and repression of origin re-replication which is an important regulation mechanism (see **Figure 1.3**) (Takisawa et al., 2000; Takeda and Dutta, 2005; Pospiech et al., 2010; Fragkos et al., 2015).

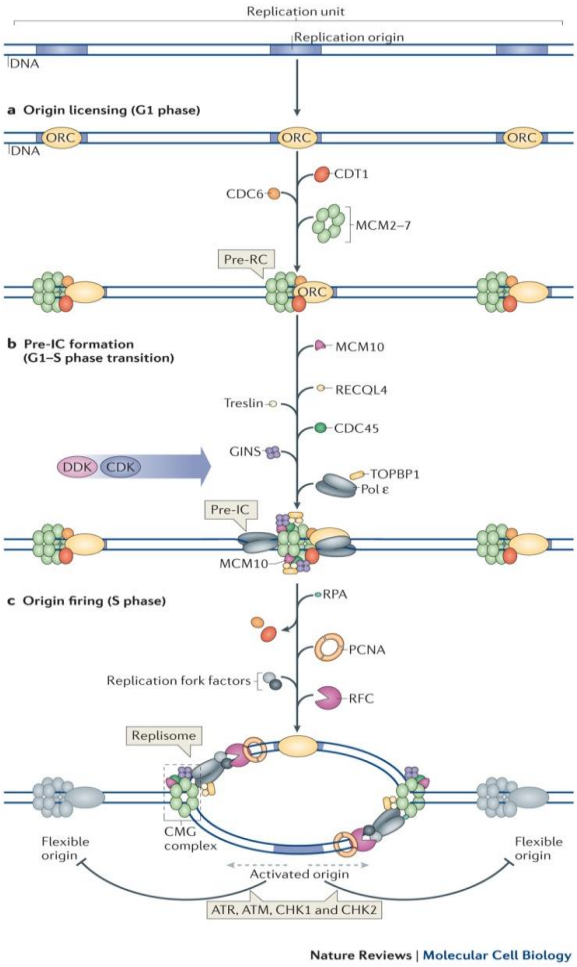


Figure 1.3. Overview of DNA replication.

A. Formation of the pre-RC complex.

In the G1 phase the proteins CDC6 and CDT1 bind the multiprotein origin recognition complex (ORC) at the origins of replication. This results in the recruitment of minichromosome maintenance proteins (MCM2-7) to the ORC complex and the formation of the prereplicative complex (pre-RC).

B. Formation of the pre-initiation complex (pre-IC).

During the G1-S transition phase the S phase cyclin dependent CDK2 and DBF4 dependent kinase CDC7 control the activation of the origins and formation of replication forks by phosphorylation of MCM10, the cell division control protein CDC45, the ATP-dependent DNA helicase Q4 (RECQL4), Treslin, GINS, DNA topoisomerase 2-binding protein (TOPBP1) and DNA polymerase ε.

C. Activation of origin.

After the assembly of the pre-IC, in the S phase the replication protein A (RPA), proliferating cell nuclear antigen (PCNA), the replication protein C (RPC) and additional replication fork proteins are recruited resulting in the formation of replication forks and origin activation. The replisome contains the CMG complex (CDC45, MCM hexamer, GINS complex) and is responsible for helicase activity and DNA unwinding. The checkpoint kinases ATR, ATM that phosphorylate the checkpoint kinases 1 and 2 (CHK1, CHK2) control and prevent the re-replication of origins. Figure was taken from Fragkos *et al*., 2015 (permission obtained).

In order to identify novel DNA replication factors and associated proteins involved in DNA replication a human cell-free system was established. In this approach template nuclei of cells that are arrested at late G1-phase are isolated. In these nuclei DNA replication can still occur but they lack active DNA replication forks. Interestingly, upon incubation of late G1 phase arrested with cytosolic extracts derived from S phase replicating cells that contain replication initiation factors DNA replication can be initiated in around 50 % of late G1-phase template nuclei (Krude 2000; Krude et al., 2009).

Using the human cell free system, the cytosolic extract from proliferating human cells was first separated by anion exchange chromatography. The obtained fractions were added to the late G1-phase arrested template nuclei. Surprisingly, in these reconstitution experiments no protein was identified. However, non-coding Y RNAs were discovered as essential factors of chromosomal DNA replication (Christov et al., 2006).

It was confirmed that Y RNAs are necessary for the initiation of DNA replication as specific knockdown of human Y RNA in cytosolic cell extracts by antisense DNA oligonucleotides resulted in the inhibition of DNA replication and cell proliferation. Intriguingly, the knockdown of Y RNAs could be rescued by the addition of any of the four human Y RNAs and DNA replication could be initiated (Christov et al., 2006).

Notably, human Y RNAs were able to initiate chromosomal DNA replication in isolated mouse cell nuclei while non-vertebrate Y RNAs were not. This suggested that the functional role of Y RNAs as essential factors in the initiation of DNA replication is conserved among vertebrates (Gardiner et al., 2009).

Further it was shown by DNA fibre fluorescence microscopy and nascent-strand analysis that Y RNAs are specifically required for the initiation of chromosomal DNA replication and establishing of new replication forks and were not required for elongation of DNA replication forks (Krude et al., 2009).

Looking at the sequence and structural requirements of Y RNAs during DNA replication mutagenesis analysis of hY1 revealed that a conserved, double-stranded 8 nt consensus sequence motif 5’-GUAGUGGG-3’ in the upper stem domain of the Y RNAs is essential for the functional role in chromosomal DNA replication. Transfection of a synthetic double-stranded RNA derived from the upper stem of hY1 alone was sufficient to initiate chromosomal DNA replication (Gardiner et al., 2009).

Using biophysical techniques such as solution state nuclear magnetic resonance (NMR) and far-UV circular dichroism (CD) spectrometry Wang *et al*. investigated whether there are any sequence and structural requirements of the conserved stem of hY1 in order to initiate chromosomal DNA replication. It was shown that mutations introduced in the conserved GUG-CAC motif resulted in a destabilization of the A-form helix and inhibition of chromosomal DNA replication (Wang et al., 2014).

Since Y RNAs are forming RoRNPs with Ro and La and Nucleolin interacts with the pyrimidine-rich internal loops of hY1 and hY3 RNAs it was tested if these proteins are required for the initiation of DNA replication. Depletion of Ro60, La and Nucleolin proteins from human cell extracts by immune precipitation did not have any effect on the initiation of DNA replication. This suggested that Ro60, La and Nucleolin were not required for the initiation of DNA replication (Langley et al., 2010).

Previous mutagenesis experiments on hY1 already showed that the conserved lower stem region of Y RNAs, where Ro60 binds, was not essential for the initiation DNA replication (Christov et al., 2006). This finding was further confirmed by reconstitution experiments which confirmed that the addition of recombinant purified Ro60, La and Nucleolin did not have any effect on DNA replication *in vitro* (Langley et al., 2010).

Later it was observed that during DNA replication all Y RNAs associate dynamically with euchromatin and replication proteins such as ORC, pre-RC and pre-IC proteins Cdt1, Mcm2 and Cdc45, but not with DNA replication fork proteins in G1 phase nuclei. These results suggested that Y RNAs might act as replication licensing factors (Zhang et al., 2011). The licensing of replication origins takes only place once per cell cycle in the G1 phase and involves the recognition of the pre-RC complex and this step allows to control that DNA is accurately replicated (Fragkos et al., 2015).

Experiments using *X. laevis* and *D. rerio* embryos revealed that Y RNAs are essential for chromosomal DNA replication after the mid-blastula transition stage in early development dependent on ORC. However, just after gastrulation a knockdown of Y RNAs by specific inhibitory morpholino oligonucleotides resulted in the rapid inhibition of DNA replication, cell cycle arrest and death of the embryos (Collart et al., 2011). This would suggest that Y RNAs are not required for the initiation of DNA replication during early development and that other factors are involved.

Recently, the nucleosome and histone deacetylation complex NuRD isolated from *X. laevis* eggs and pre-MBT embryos was identified to be responsible for the initiation of DNA replication in early embryogenesis before MBT in the absence of non-coding Y RNAs.

The inactivation of xNuRD led to an inhibition of DNA replication, developmental arrest and embryonic lethality. However, it was shown that in contrast to xNuRD complexes human NuRD complexes were not able to initiate DNA replication. Taken together, it was suggested that the xNuRD complex plays an important role in the regulation of DNA replication through chromatin remodelling during early development. After MBT Y RNAs might also be involved in chromatin remodelling in order to activate replication forks and initiate DNA replication (Christov et al., 2018).

Another recent study looked more into the detail of the molecular mechanism of Y RNA function and localization during cell cycle. Using proliferating cells derived from human cancer and non-cancerous cell lines the Y RNAs were quantified in the soluble and chromatin-associated fractions by qRT-PCR. Around one million soluble Y RNA molecules per cell were quantified and among all Y RNAs it was found that Y3 was the most abundant. Interestingly, there was no significant difference in Y RNA accumulation between proliferating cancerous and non-cancerous cells. However, in quiescent cells the level of Y RNAs was decreased, whereas in proliferating cells the Y RNA level was upregulated which is consistent to their role in DNA replication (Kheir and Krude, 2017).

It was shown by fluorescence labelling that Y RNAs preferentially associated with S phase chromatin compared to G1 phase chromatin. Furthermore, it was demonstrated that Y RNAs are required for the dynamic interaction with the ORC complex and initiation of DNA replication. Interestingly, co-localization experiments of human Y RNAs revealed that all Y RNAs, except Y5 preferentially interacted with the ORC complex. Y5 RNA was primarily associated with Nucleolin and was mostly localized in the nucleolus. Since Y5 was previously shown to interact with the ribosomal protein L5 it was suggested that Y5 might also be important for other nuclear processes apart from the initiation of DNA replication. It was hypothesized that another possible function could be in ribosome biogenesis (Kheir and Krude, 2017).

1.8.2 Y RNAs in cancer

One study reported that all human Y RNAs were significantly upregulated (between 4- and 13-fold) in extracts derived from human solid tumours (n=42) compared to corresponding normal tissues (n=24) and cultured cells using quantitative RT-PCR. In this study six different tissue types which included urinary bladder, cervix, colon, kidney, lung and prostate were analysed and compared to normal tissue. Among all the Y RNAs, hY1 and hY3 were most significantly upregulated in the solid tumour samples (carcinomas and adenocarcinomas) including all analysed different tissue types (Christov et al., 2008).

In the next step it was investigated if Y RNAs are required for cell proliferation. When siRNAs against hY1 and hY3 were transfected into several human cell lines a significant reduction in cell proliferation was observed (Christov et al., 2008).

Recently it was shown that in tumours Y RNAs were upregulated whereas in quiescent normal tissues the level of Y RNAs was decreased. Hence, it was proposed that the expression of Y RNAs could be used as a potential cell proliferation marker instead of a cancer biomarker (Kheir and Krude, 2017).

In recent years several studies demonstrated that the expression of Y RNAs is altered in cancer cells. Using quantitative real time PCR it was observed that all Y RNAs are downregulated in urinary bladder cancer (n=88) compared to normal tissues (n=30) (Tolkach et al., 2017). In the most recent study the same group found that Y RNAs are also downregulated in prostate cancer patients (n=56) compared to normal tissues (n=36). This downregulation of Y RNAs in urinary bladder cancer and prostate cancer was controversial to the study performed by Christov *et al*. who saw an upregulation of Y RNAs in tumour samples. These contradictory results could be either explained by technical issues or the fact that Y RNA expression is different among various cancer types (Christov et al., 2008; Tolkach et al., 2018). However, the exact functions and molecular mechanisms of Y RNAs in cancer remains obscure and needs further investigation.

1.9 Y RNA derived small RNAs

Y RNA-derived sRNA (YsRNA) fragments were first described in apoptotic cells by Rutjes *et al*. in 1999. It was shown that during apoptosis human Y RNAs are rapidly and specifically cleaved in a caspase-dependent manner generating shorter YsRNA fragments between 22-25 nt and a larger fragment of 31 nt (Rutjes et al., 1999a). In this study it was suggested that the nucleases that cleave Y RNAs are caspase-dependent nucleases since Y RNA cleavage was similar to the cleavage of the U1-70K which was shown to be caspase-3-dependent (Casciola-Rosen et al., 1994; Rutjes et al., 1999a). Results from immunoprecipitation experiments showed that YsRNAs are bound to Ro60 and La. The 31 nt fragment associated with both Ro60 and La, whereas the shorter fragments only bound to Ro60 (Rutjes et al., 1999b). These YsRNAs were assumed to be degradation products for a long time and it was technically difficult to work with these short sRNA fragments (Hall et al., 2013). However, little was known about their biological role and mechanisms.

Next generation sequencing from precursor B cells derived from acute leukaemia patients and normal CD34+ cells revealed that sRNA fragments derived from the 3' end of hY3 and hY5 Y RNAs were highly enriched (Schotte et al., 2009). However, these sRNA fragments were initially mis-annotated as miRNAs and named miR-1975 (derived from the 3' end of hY5) and miR-1979 (derived from the 3' end of hY3). Meiri *et al*. also identified them in several solid tumours while they realized that the sRNAs are derived from Y RNAs. Performing luciferase reporter assays they showed that these sRNA fragments had no gene silencing activity which suggested that miR-1975 and miR-1979 were not miRNAs (Meiri et al., 2010). Consequently, miR-1975 and miR-1979 were removed from miRBase and annotated as Y RNA-derived sRNAs.

It was also shown that YsRNAs are derived from the 3' and 5' ends of Y RNAs and that the biogenesis of Y RNA derived sRNAs is independent of the canonical miRNA pathway (Nicolas et al., 2012). Y RNA fragments were found in both cancerous and non-cancerous as well as stress-induced and unstressed cells. The levels of hY5-derived sRNAs in non-stressed cells are similar to the abundance of miRNAs. Northern blot experiments revealed that the levels of Y RNA fragments derived from the 3'- and 5' end of human Y5 were significantly increased during apoptosis. Furthermore, immunoprecipitation followed by a pull down assay and anion exchange chromatography suggested that hY5 derived sRNAs are not in a complex with Argonaute 2 (Nicolas et al., 2012). This indicates that YsRNAs are not directly loaded into the RNA induced silencing complex (RISC).

Therefore, YsRNAs are not involved in miRNA biogenesis pathways and Y RNA derived sRNAs are not miRNAs (Nicolas et al., 2012).

YsRNAs have also been identified in brain, retina, and other mammalian tissue. Further, it was shown that YsRNAs were induced during viral infection (Cui et al., 2010; Qi et al., 2010). Recently, YsRNA fragments have been identified in human serum and plasma as circulating components of larger complexes (100-300 kDa). These circulating Y RNA fragments were found cell-free and not included in exosomes or microvesicles (Dhahbi et al., 2013). Moreover, YsRNAs are upregulated in the serum of breast cancer patients. Hence, YsRNAs circulating in human serum and plasma might be involved in cancer and therefore be potential biomarkers (Dhahbi et al., 2014).

RNA sequencing experiments showed that around 95% of serum derived sRNA sequencing reads mapped to hY4 RNA and its pseudogenes, whereas only a small number of Y RNAs mapped to hY1, hY3 and hY5 RNAs (Dhahbi et al., 2013). The hY4 RNAs circulating in the serum were cleaved into 5’ end Y RNA fragments at different positions and distinct proportions in the predicted internal loop of the full-length hY4 RNAs (Dhahbi et al., 2013). Interestingly, in human serum the 5’ end derived Y RNA fragments were roughly 20 times more abundant than the 3’ end derived Y RNA fragments (Dhahbi et al., 2014). The levels of 5’ end derived Y RNA fragments were significantly increased in human serum compared to mouse serum (Dhahbi et al., 2013). It needs to be further investigated why the composition of circulating sRNAs is different in human and mouse sera. Also, Y RNA fragments have been detected in extracellular vesicles (EVs) that were released from human semen and immune cells (Nolte-'t Hoen et al., 2012; Vojtech et al., 2014).

A recent study investigated the processing and transfer of hY5 RNA fragments within EVs derived from primary and cancer cell lines. Intriguingly, hY5 RNAs were highly abundant and enriched in EVs. Shorter fragments (23, 29 and 31 nt) derived from the 5’ and 3’ end of hY5 were also detected in EVs (Chakrabortty et al., 2015). In the same study hY5 RNA fragmentation occurred in the EVs instead of inside the cells. The 5’ end derived hY5 RNA fragments contained part of the conserved 8 nt motif (5’-GUAGUGGG-3’) which is involved in the initiation of chromosomal replication (Gardiner et al., 2009). The same motif was shown to be important for the processing of full-length hY5 RNAs into smaller Y RNA fragments (Chakrabortty et al., 2015).

Interestingly, the exposure of human healthy cells with EVs derived from cancer cells resulted in rapid cell death. However, cancer cells did not show any phenotype when they were treated with EVs from normal or cancer cells. A synthetic version of the 23 nt- and 31 nt 5' end derived hY5 fragment triggers cell death in primary cells in a dose-dependent manner through an apoptotic pathway. Further, it was observed that the conserved 8 nt motif in the 5' end derived Y RNA fragments was required for the induction of cell death (Chakrabortty et al., 2015).

In conclusion, there seems to be a correlation between 5' end derived Y RNA fragments and cancer, since 5' end derived Y RNA fragments released from cancer cell EVs trigger rapid cell death in normal cells. It was hypothesized that the 5' Y RNA fragment induced cell death might help tumours to invade tissues and to establish a microenvironment (Chakrabortty et al., 2015).

1.10 Functions of Y RNA derived small RNAs

1.10.1 Y RNA derived small RNAs involved in coronary artery disease

Coronary artery disease (CAD) is one of the most common cardiovascular diseases and affects millions of people worldwide. At present it is the one of the leading causes of death. CAD is caused by atherosclerosis of heart arteries and it was shown that macrophage apoptosis is associated with atherogenesis pathogenesis. In early stages of atherogenesis the accumulation of lipoproteins, mainly low-density lipoprotein (LDL) results in the activation of immune response and recruitment of macrophages. By endocytosis the macrophages ingest lipids and lipoproteins and become so called foam cells. The foam cell formation results in the activation of cytokines and apoptosis followed by a rapid phagocytic clearance of apoptotic foam cells and decrease of atherosclerotic lesion progression. However, when macrophage apoptosis and clearance of apoptotic foam cells are not tightly regulated the inflammatory response results in atherosclerotic lesions and necrotic plague formation.

Interestingly, next generation sequencing of human and mouse apoptotic macrophages revealed that the expression of Y RNA derived sRNAs is induced after treatment with pro-apoptotic or atherogenic stimuli.

Further, in qRT-PCR experiments the levels of YsRNAs was significantly increased in mouse models for atherosclerosis as well as the serum of 263 patients with coronary artery disease (CAD) compared to 514 healthy controls. Among of all Y RNAs, the hY1 5' end derived sRNA fragment was the most upregulated Y RNA derived sRNA in the serum CAD patient samples. Statistical analysis indicated that circulating Y RNA derived sRNA fragments might be used as novel diagnostic biomarkers for CAD and other related cardiovascular diseases (Hizir et al., 2017).

Interestingly, a 31 nt 5' end derived hY4 derived sRNA was reported to be associated with platelet function in patients with acute coronary disease and have been shown highly abundant in plasma and enriched in exosomes (Dhahbi et al., 2013). Using next generation sequencing and quantitative real time PCR with molecular beacons a 31/32 nt 5' end derived hY4 fragment was detected in saliva of healthy persons suggesting that this 5' end derived hY4 sRNA could be used as a potential biomarker to enhance diagnosis of coronary diseases (Ishikawa et al., 2017).

1.10.2 Y RNA fragment involved in cardioprotection

Recently, it was found that Y RNA fragments are highly enriched in extracellular vesicles (EV) derived from cardiosphere derived cells (CDC) that were shown to reduce myocardial infarct size. CDCs secrete EVs and are involved in inflammatory response and regeneration of myocardial infarcts through polarization of macrophages. Among all the Y RNAs the 5' end derived hY4 fragment named EV-YF1 was the most abundant in CDC EVs. Transfection of fluorescently labelled EV-Y1 into macrophages induced an upregulation of interleukin IL-10 gene expression and led to cardioprotection and a reduction of infarct size (Cambier et al., 2017).

1.10.3 Y RNA derived sRNAs regulate cell death and inflammation in monocytes/macrophages

Intriguingly, a recent study demonstrated that Y RNA derived sRNAs regulate apoptosis and inflammatory response in monocytes/macrophages. When monocytes/macrophages were treated with pro-atherogenic stimuli such as oxLDL or palmitic acid (PA) caspase-dependent apoptosis and the NFκβ signalling pathway was activated.

Interestingly, the co-transfection of 2'-O Me-RNA oligonucleotides that were antisense to YsRNAs and the knockdown of YsRNAs using siRNAs resulted both in the downregulation of apoptosis in monocytes/macrophages (Hizir et al., 2017).

In order to investigate if extracellular YsRNA/Ro60 complex is involved in apoptosis and inflammation in monocytes/macrophages Y RNA/Ro60 and YsRNA/Ro60 were immunopurified either from treated cells or from untreated control cells. Monocytes were treated with Y RNA/Ro60 or YsRNA/Ro60 complexes and apoptosis was induced after 12 h.

When the monocytes were cotransfected with antisense YsRNAs apoptosis was inhibited, indicating that the extracellular YsRNA/Ro60 activates apoptosis and NFκβ signalling. The transfection with YsRNAs or the full-length Y RNA/Ro60 complex did not show any activation of apoptosis. From these experiments it can be concluded that extracellular YsRNAs need to be in complex with Ro60 in order to be able to enter the cells and induce apoptosis and inflammation (Hizir et al., 2017).

In previous experiments it was shown that Ro60 is associated with endogenous retroelements and plays an important role in the regulation of inflammatory gene expression. When Ro60 was knocked down the expression of endogenous Alu transcripts and IFN response related genes were upregulated. Since the upregulation of Alu transcripts activated the secretion of cytokines it was tested whether the inflammatory response was activated through Toll-like receptors (TLR). Treatment of cells with chloroquine that inhibits the signaling pathway of TLR3, 7, 8 and 9 and IRS954-thioate that blocks TLR7/9 revealed that the cytokine secretion was TLR7 dependent (Hung et al., 2015).

In order to test if TLRs are also involved in the YsRNA induced apoptosis and inflammation in monocytes/macrophages the human and mouse monocytes were treated with chloroquine and IRS954, subjected to flow cytometry and validated by western blot analysis. It was shown that chloroquine and IRS954 inhibited the YsRNA induced activation of caspase-dependent apoptosis and NFκβ inflammatory response. These results indicate that intracellular and extracellular YsRNAs stimulate TLR7 to mediate apoptosis and inflammation in both human and mouse monocytes/macrophages. The sequence and secondary structure of Y RNAs is highly evolutionary conserved among vertebrates (Pruijn et al., 1993). Interestingly, the conserved function of YsRNAs between human and mouse monocytes also indicates that the function of YsRNAs might be evolutionary conserved and might underlie a similar mechanism (Hung et al., 2015).

1.10.4 Human Y4 RNA fragment can act as a guide for tRNase Z

Next generation sequencing of plasma derived sRNAs from three healthy persons and three myeloma patients revealed that a 31 nt 5' end derived hY4 fragment was significantly more abundant in myeloma patients compared to the healthy control persons. Intriguingly, this hY4-derived sRNA can adopt the structure of a 5'-tRNA half and associate with a target RNA in order to form a pre-tRNA-like structure. This structure is then recognized and cleaved by tRNase Z (Ninomiya et al., 2015).

Interestingly, recently a new class of bacterial Y RNAs named Y RNA-like A (YrlA) RNAs was discovered. By crystal structure of *S. thyphimurium* YrlA it was shown that this bacterial Y RNA is able to fold like a tRNA (Wang et al., 2018). By Chen *et al*., it was already previously observed that bacterial Y RNAs can mimic tRNAs and can be processed like tRNAs. Additionally, they could demonstrate that these Y RNAs had nucleotide modifications which allowed the recognition of ribonucleases involved in tRNA cleavage (Chen et.,2014).

1.10.5 PTBP1

Notably, during apoptosis Y RNAs are cleaved into smaller fragments in a caspase dependent manner (Casciola-Rosen et al., 1994; Rutjes et al., 1999a; Chen et al., 2014). Next generation sequencing of Epstein-Barr virus (EBV) positive lymphoma cells revealed a new class of small RNAs named AGO-taxis small RNAs (ASRs) that were specifically bound to Ago1. These ASRs were highly upregulated during lytic EBV infection and it was suggested that they are involved in apoptosis. Further it was observed in siRNA mediated knockdown experiments that the biogenesis of ASRs was independent of Drosha, Dicer and RNase L. ASRs were shown to have gene silencing activity and CLASH (crosslinking, ligation, and sequencing of hybrids) data revealed that ASRs primarily bound the coding sequence of the identified targets (Yamakawa et al., 2014).

Recently, the same group reported that Y RNAs might be the precursors of ASRs and that ASRs are generated by Y RNA cleavage during apoptosis. Cell fractionation experiments using saturated ammonium sulfate (SAS), ion exchange chromatography (IEC) and gel filtration chromatography resulted in the identification of a cytoplasmic endoribonuclease and its inhibitor PTBP1 which is also known as hnRNP1 and binds to the loop region of Y RNAs (Fabini et al., 2000; Ogata et al.,

2018). It was shown that PTBP1 inhibited Y RNA cleavage by blocking the access of endoribonuclease (s) to the central loop.

During apoptosis, PTBP1 gets cleaved by caspase 3 which results in Y RNA cleavage and production of ASRs. However, the endoribonuclease involved in the generation of the 3' end derived ASR could not be identified and further investigation will be needed (Ogata et al., 2018).

1.10.6 RNase L

During viral infection dsRNA triggers the activation of the innate immune response through type-I interferons (IFNs) which result in the activation of 2' 5' oligoadenylate synthetase genes (OAS) (Hovanessian et al., 1977). These OAS genes function as pathogen pattern recognition receptors (PPRs) in order to initiate interferon induced antiviral response. The activation of OAS genes by viral dsRNA leads to the synthesis of small molecules named 2-5 A oligoadenylates in the presence of ATP (Silverman, 2007; Chakrabarti et al., 2015).

The endoribonuclease RNase L was shown to bind 2-5A with low affinity at subnanomolar levels. The binding of 2-5A leads to the formation of catalytically active homodimers. RNase L only cleaves ssRNAs preferentially after U-rich sequences but not dsRNA molecules (Dong and Silverman, 1995). RNase L belongs to the family of metal ion independent ribonucleases (Cooper et al., 2014). RNase L preferentially cleaves ssRNAs after UpAp and UpUp dinucleotides (Wreschner et al., 1981) and has sequence specificity for the consensus sequence UN^N (N = A, C, G, U). It was observed that RNase L produces 5'-hydroxyl and 2', 3' cyclic phosphate termini at the 3' end of cleaved RNAs (Han et al., 2014).

In order to be able to sequence small RNAs with 2' 3' cyclic phosphates Donovan *et al*. used a cDNA library construction method based on RtcB ligase which allows the ligation of RNAs with 2' 3' cyclic phosphates to the adapter sequence (Tanaka et al., 2011).

Next generation sequencing of HeLa cells treated with poly (I:C) which mimics viral infection and untreated cells revealed that tRNAs and Y RNAs were specifically cleaved at UN^N position. Y RNA cleavage primarily occurred at the 5' end of Y RNAs between the positions 24-32 nt. However, hY4 RNA was exceptional and was cleaved

after CA^G and it was suggested that hY4 RNA might be cleaved by another ribonuclease.

For the most abundant tRNAs in the dataset tRNA-His, tRNA-Gln and tRNA-Pro it was shown that the tRNAs were cleaved specifically in the anticodon stem loop region during antiviral response. Further overexpression experiments of wild type RNase L and RNase L mutant H672N revealed that tRNAs and Y RNAs were only cleaved in cells in which wild type RNase L was overexpressed. Treatment of cells with 2-5A also resulted in tRNA and Y RNA cleavage. Using human RNase L $^{-/-}$ cells tRNA and Y RNA cleavage could only be detected by Northern blotting in RNase L wild type cells. These results indicated that RNase L is directly responsible for tRNA and Y RNA cleavage (Donovan et al., 2017).

Interestingly, it was reported that RNase L contributes to the early arrest of translation during dsRNA response. Previously it was shown that the protein kinase PKR is activated by dsRNA during antiviral response and is involved in translational inhibition through the phosphorylation of the translation initiation factor eIF2α (Smith et al., 2005). By the analysis of nascent protein synthesis in wild type and RNase L $^{-/-}$ cells it was demonstrated that RNase L is critical to inhibit protein synthesis. Intriguingly, it was shown that translational arrest occurs rapidly, whereas components of the translation machinery including rRNAs or mRNAs are degraded much later. It was proposed that tRNA and Y RNA derived fragments might be involved in a signalling like mechanism with RNase L and regulate translational inhibition. However the exact mechanism how Y RNA and tRNA fragments are involved in this process is not clear (Donovan et al., 2017).

1.11 Aims and objectives of the thesis

Although YsRNAs were initially discovered in apoptotic cells almost two decades ago and YsRNAs have been described in many different biological aspects during the last two years, the exact biogenesis and mechanisms of YsRNAs are still remain poorly characterized and further investigation is needed. It is currently unknown, what role YsRNAs play in normal cells and how they function mechanistically during apoptosis, stress, or cancer.

In this thesis, I took two approaches in order to understand the mechanisms underlying the biogenesis of YsRNAs:

As Y RNAs are highly evolutionary conserved in vertebrates in the first two result chapters of this thesis I investigated if the nucleotide sequence or conserved structural features contribute to the size-specific Y RNA cleavage.

The aim of the first two thesis chapters was to establish sequence and structural requirements for Y RNA cleavage into YsRNAs in order to understand how Y RNAs are processed during stress response. Within this investigation I aimed to address the question if the nucleotide sequence or the secondary structure is more important in respect to determine where the Y RNA cleavage occurs.
In the first result chapter I aimed to further validate and analyse the next generation sequencing results from a high throughput mutagenesis approach on the 3' end of the human Y5 RNA (**Figure 1.4, A**). The cDNA libraries of the 3' end hY5 RNA mutants were generated by a previous PhD student Carly Turnbull (Turnbull, 2013).

I took over this project and my aim was to experimentally validate the sequencing results of the 3' end hY5 RNA mutagenesis approach by generating individual hY5 RNA mutants and testing their expression and cleavage pattern by Northern blot analysis.

Additionally, the aim of the first thesis chapter was to explore if there was a minimal nucleotide requirement for Y RNA cleavage to occur. In particular, in subsequent experiments I investigated if the sequence of the internal L2a loop (**Figure 1.4, B**) was important for 3' end hY5 RNA cleavage.

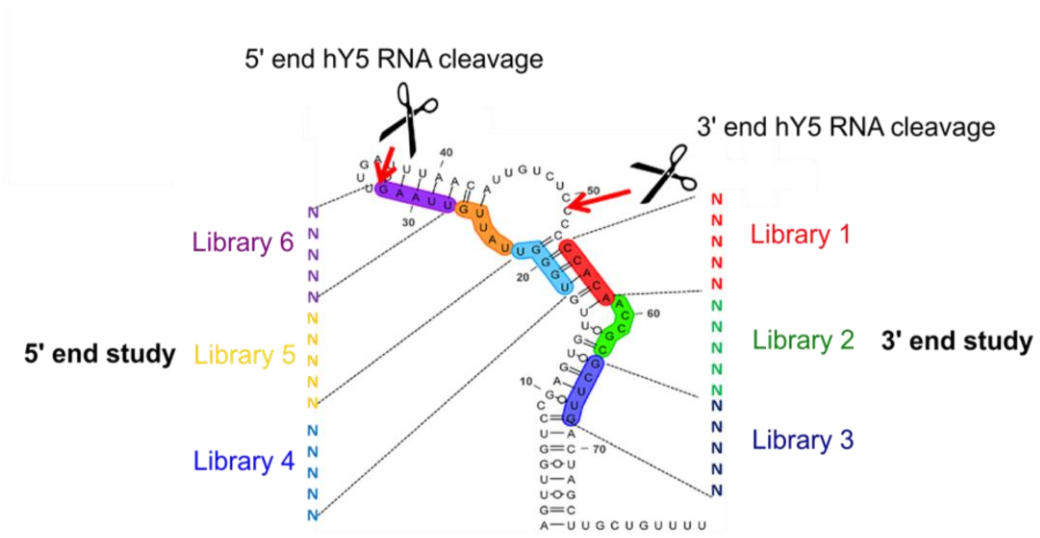
Due to advances of next generation sequencing technologies it became apparent that 5’ end derived hY5 RNA fragments are functional.

Hence, in the second result chapter of this thesis my objectives were to design and to perform a high throughput mutagenesis approach on the 5’ end of hY5 RNA.

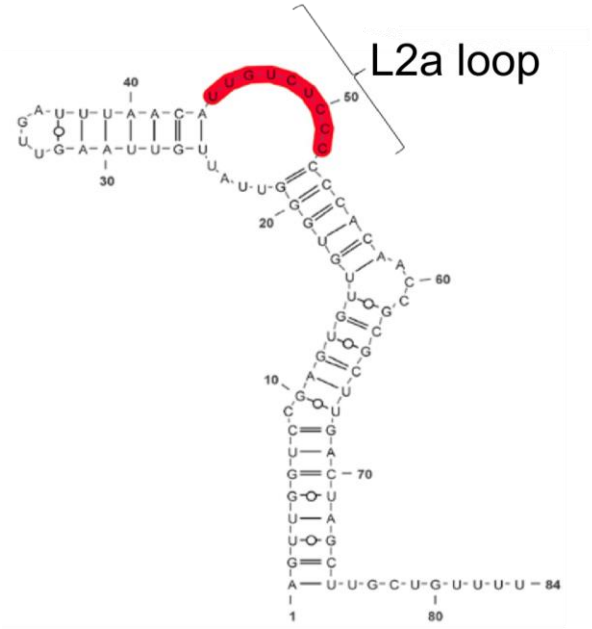
My goal was to validate and further analyse the sequencing results by cloning and testing the expression levels of individual hY5 RNA mutants by Northern blot analysis. In addition, my aim was to investigate if the sequence and/or structure of the internal loop L2b (**Figure 1.4, C**) was involved in Y RNA cleavage.

In the last chapter of this thesis I tried to use a biochemical approach in order to understand which protein(s) or ribonuclease(s) are involved in Y RNA cleavage. In particular, my goal was to examine of role of Ro60 and RNase L on Y RNA cleavage (**Figure 1.4, D**) in more detail.

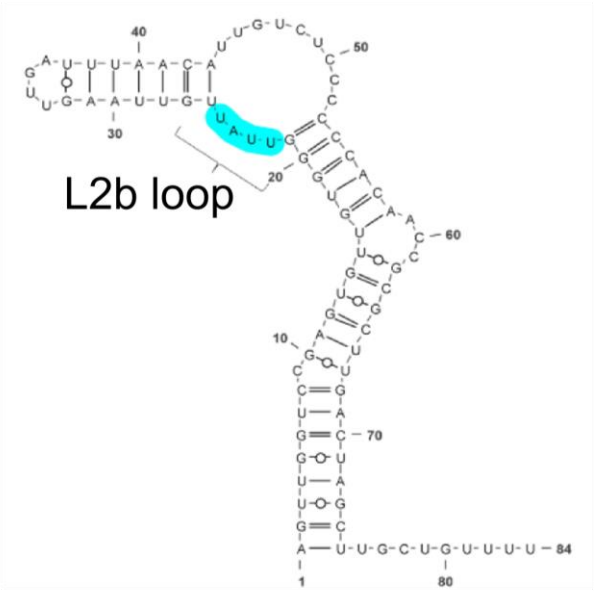
A


5' end hY5 RNA cleavage
3' end hY5 RNA cleavage
Library 6
Library 5
Library 4
5' end study
Library 1
Library 2
Library 3
3' end study

B


L2a loop

C


L2b loop

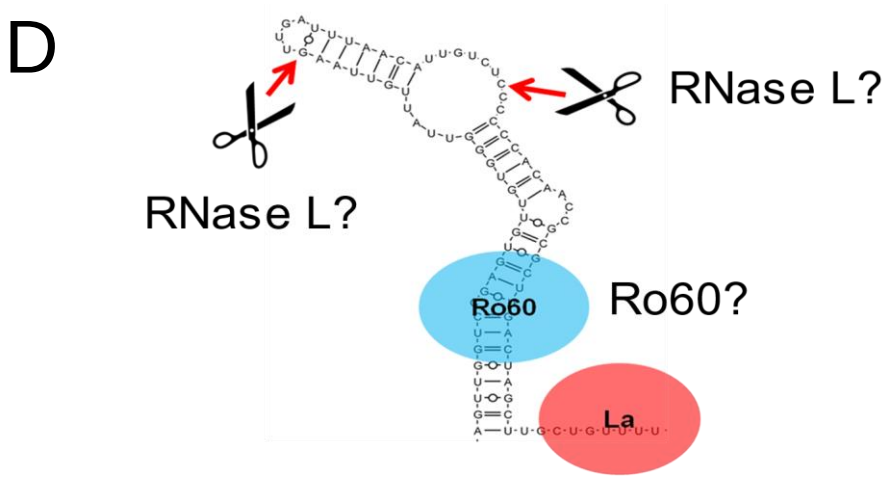


Figure 1.4. Overview of aims and objectives performed in this thesis.

This figure gives an overview about the aims and objectives of this thesis.

A. High throughput mutagenesis approach on the 3‘ and 5’ end of hY5 RNA.

The secondary structure of hY5 RNA was predicted with RNAfold and visualized using Varna. The human Y5 RNA has a characteristic stem loop structure. Three different regions near the cleavage sites at the 3’/5’ end of the hY5 RNA were selected for random mutagenesis. The cleavage sites on the 3’ and 5’ end of hY5 RNA are indicated with red arrows and unknown RNases involved in hY5 RNA cleavage are depicted as scissors. In each of these regions close to the 3’ and 5’ end cleavage sites 5N substitution mutations were introduced resulting in 1024 (4^5) possible combinations for each library. Each library is shown in different colors. For the 3’ end high throughput mutagenesis analysis library 1 (red), library 2 (green) and library 3 (dark blue) were generated. The 5’ end high throughput mutagenesis involved the generation of library 4 (blue), library 5 (yellow), library 6 (purple).

B. Mutagenesis of the hY5 RNA internal loop L2a.

The secondary structure of hY5 RNA was predicted with RNAfold and visualized using Varna. The human Y5 RNA has a characteristic stem loop structure. One aim of the thesis was to mutate the internal loop 2a (shown in red) in order to investigate if the length or nucleotide sequence of this internal loop is important for 3’ end hY5 RNA cleavage.

C. Mutagenesis of the hY5 RNA internal loop L2b.

The secondary structure of hY5 RNA was predicted with RNAfold and visualized using Varna. The human Y5 RNA has a characteristic stem loop structure. One aim of the thesis was to mutate the internal loop L2b (shown in turquoise) in order to investigate if the sequence or secondary structure of the internal loop L2b is important for human Y RNA cleavage length or nucleotide sequence of this internal loop is important for 3’ end hY5 RNA cleavage

D. Investigation of which proteins/ribonucleases are involved in Y RNA cleavage.

In the last chapter it was examined in more detail if Ro60 and RNase L are involved in Y RNA cleavage.

Chapter 2 Materials and Methods

2.1 Cell culture

2.1.1 Cell lines and cell culture media

NIH 3T3 cells were purchased from the European Collection of Authenticated Cell Cultures (ECACC) and were cultured in Dulbecco's modified Eagle's medium (DMEM) (Gibco, 10938) supplemented with 10% heat inactivated fetal bovine serum (FBS) (Gibco, 10500), 1% L-Glutamine (200 mM) (Gibco, 25030) and 1% Penicillin-Streptomycin (10000 U/ml) (Gibco, 15140).

HeLa cells and MCF7 cells were kindly provided by the laboratory of Dylan Edwards. The cell lines were subjected to cell line authentication and mycoplasma testing by ECACC (European Collection of Authenticated Cell Cultures). HeLa cells were cultivated in Eagle's Minimum Essential Medium (M2279, Sigma), supplemented with 10% heat inactivated fetal bovine serum (FBS), 1% MEM Non-Essential Amino Acid Solution (100x) (Gibco, 11140), 1% L-Glutamine (200 mM) and 1% Penicillin-Streptomycin (10000 U/ml) (Gibco, 15140).

MCF7 cells were cultured in low glucose DMEM + GlutaMAX medium (Gibco, 21885) that was supplemented with 10% heat inactivated fetal bovine serum (FBS) (Gibco, 10500, 1% MEM Non-Essential Amino Acid Solution (100x) and 1% Penicillin-Streptomycin (10000 U/ml) (Gibco, 15140).

The wild type and Ro60 $^{-/-}$ mouse embryonic stem cells (mES) were a kind gift from Prof. Sandra Wolin (NCI, US) (Chen *et a*l., 2003). The mES cells were cultured in Knockout DMEM (Gibco, 10829), 15% ES cell FBS (Gibco, 10439), 1% L-Glutamine (200 mM) (Gibco, 25030), 1% MEM Non-Essential Amino Acid Solution (100x) (Gibco, 11140), 100 U/ml ESGRO leukemia inhibitory factor (Millipore, ESG 1107) and 0.1 mM 2-Mercaptoethanol (Sigma, M6250).

The wild type and RNase L $^{-/-}$ mouse embryonic fibroblast cells (mEFs) were kindly provided by Prof. Robert Silverman. The RNase L $^{-/-}$ mEF cells were generated by the insertion of a neomycin resistance gene into the 5' end of RNase L gene (Zhou et al., 1997).

The human wild type and CRISPR RNase L $^{-/-}$ A549 lung cancer cells were a kind gift of Prof. Susan Weiss (Cleaveland Clinic, US) (Li et al., 2016).

The mouse and human wild type and RNase L $^{-/-}$ cell lines were cultivated in RPMI 1640 medium (Gibco, 11975-001) supplemented with 1% L-Glutamine (200 mM) (Gibco, 25030) and 1% Penicillin-Streptomycin (10000 U/ml) (Gibco, 15140).

2.1.2 Sub-culturing and cell culture conditions

All cell lines used in this thesis were cultivated in a humidified CO_2 incubator under standard conditions at 5% CO_2 and 37°C. All cell culture work was performed in a microbiological safety cabinet using aseptic technique.

For passaging the cells, the cell culture media was removed and the cells were once washed with PBS (Gibco, 10010). For NIH 3T3, HeLa and MCF7 cells 0.25% Trypsin-EDTA (Gibco, 25200) was added to the flask and pipetted gently to cover all the cells. The flask was placed in the incubator for 2-5 min at 37°C. After incubation the flask was gently tapped a few times to ensure all the cells were detached. The Trypsin-EDTA was inactivated by adding an equal volume of the appropriate cell culture media to the flask. The cell suspension was centrifuged for 5 min at 1,500 rpm. The supernatant was carefully removed and the cells were resuspended in 10 ml fresh cell culture media. The cell number was counted using a haemocytometer. For that 100 µl of cell suspension was added to 400 µl of 0.4% Tryphan blue and mixed gently. Tryphan blue was used to calculate the cell viability because it can enter dead cells and live cells do not take up the dye. To calculate the number of viable cells 20 µl of the cell suspension was applied to the haemocytometer and the number of cells was counted accordingly.
The cell suspension was diluted at the appropriate seeding density with fresh cell culture media and transferred to a new flask containing the required volume of media. The wild type and Ro60 $^{-/-}$ mouse embryonic stem cells were passaged using 0.05% Trypsin-EDTA.

2.1.3 Freezing cells

For freezing the cells were trypsinized and centrifuged for 5 min at 800 rpm. The cell pellet was resuspended in cell freezing medium containing 80% media, 10% FBS and 10 % dimethyl sulfoxide (DMSO) (Sigma, D2650). In each cryovial 1 ml of cell suspension ranging in concentration from 2-5 x 10^6 cells/ml was added and placed into an isopropanol bath at -80°C overnight. The next day the cryovials were placed into a liquid nitrogen Dewar for long term storage.

2.1.4 Transfection of NIH 3T3 cells with plasmid DNA

The transfections were carried out in 6-well plates (Sarstedt). The cells were seeded at 200,000 cells/well the day before transfection using cell culture media without antibiotics. The cells were transfected using Fugene 6 (Promega, E2691) according to the manufacturer's manual. For the transfection of a 6-well plate 60 µl of Fugene 6 was diluted in serum-free OptiMEM I reduced-serum medium (Gibco, 51985), gently mixed and incubated for 5 min at room temperature. The recommended ratio between transfection reagent and DNA ratio for NIH 3T3 cells was 3:1, which means for each 1 µg of DNA 3 µl of Fugene 6 reagent was used. For one 6 well plate 20 µg of plasmid DNA was added to the transfection mix, pipetted several times and incubated for 15 min at room temperature. During incubation the cell culture media of the 6 well plate was replaced with fresh media without antibiotics and 150 µl of the transfection mixture was added drop wise to the cells. For each transfection experiment appropriate control experiments such as transfection with mock or empty vector were conducted.

2.1.5 Transfection of MCF7 and mES wild type and Ro60 $^{-/-}$ cells with poly (I:C)

Polyinosinic: polycytidylic acid potassium salt (poly (I:C)) (Sigma, P9582) was prepared in sterilized filtered water at a concentration of 2.5 mg/ml and was used in the experiments at a final concentration of 10 µg/ml.

For the poly (I:C) transfection experiments MCF7 and mES wild type and Ro60 $^{-/-}$ cells were seeded in 6-well plates at a cell density of 200,000 cells/well in cell culture media without antibiotics the day before transfection.
The cells were transfected using Lipofectamine ^{TM} 2000 transfection reagent (Thermo Fisher Scientific, 11668027). For one 6-well plate 32 µl Lipofectamine was gently mixed with 4 ml Opti-MEM reduced-serum medium. In another 15 ml falcon tube 32 µl poly (I:C) were mixed with 4 ml Opti-MEM and incubated for 5 min at room temperature. After the incubation the Lipofectamine-Opti-MEM mixture was added to the tube containing poly (I:C)-Opti-MEM and incubated for 20 min.
The cell culture media was removed, the cells were washed three times with PBS and fresh OptiMEM was added to the 6-well plate. After 5 h of transfection the Opti-MEM was replaced with fresh cell culture media. The cells were treated with poly (I:C) in a total of 8 h and cells were harvested after 3 h for RNA extraction.

2.1.6 Treatment of NIH 3T3 cells with staurosporine

NIH 3T3 cells were seeded at 200,000 cells/ml in a 6-well plate one day before transfection. The cells were transfected with plasmid DNA using Fugene 6 or Lipofectamine and after 24 h the cells were treated with staurosporine.
Staurosporine (Cell Signalling Technology, #9953) was prepared in DMSO at a concentration of 100 µM. The cells were treated with staurosporine at a final concentration of 1 µM for 4-8 h. After the treatment the cells were harvested for RNA extraction.

2.2 Protein extraction

For cell lysis and protein extraction of wild type and Ro60 $^{-/-}$ mouse embryonic stem cells RIPA buffer (Sigma, R0278) was used according to the manufacturer's instructions. In order to prevent protein degradation a protease inhibitor cocktail was added to the lysis buffer. After removal of the growth medium the cells were washed with cold Phosphate Buffered Saline (PBS) (Gibco, 14190). Then 300 µl of RIPA lysis buffer was added per well of a 6-well plate and incubated in the cold room under rotation for 30 min.

After the incubation the cells were removed with a cell scraper and vigorous pipetting and incubated on ice for another 5 min. The protein lysate was clarified by centrifugation at 12,000 rpm for 20 min at 4°C. The lysate was frozen in liquid nitrogen and stored at -80°C.

The protein concentration was determined in triplicates using Pierce bicinchoninic acid (BCA) protein assay kit (Thermo Scientific, 23227) following the manufacturer's protocol. The standard curve was prepared using Bovine Serum Albumin (BSA) of known concentrations. The protein concentration was calculated from the standard curve.

2.3 RNA extraction

For the work with RNA all the surfaces, pipettes and gloves were cleaned with 10% Distel (VWR) as a surface detergent in order to avoid RNA degradation by RNases. For preparing RNase-free solutions analytical reagent grade water (Fisher Scientific) was autoclaved and filter sterilized using a 0.2 µm filter (Sartorius).

Total RNA was extracted using TRI Reagent (Thermo Fisher Scientific, AM9738) according to the manufacturer's instructions.

The cell culture media of the 6-well plate was removed and the cells were washed once with PBS. When cells were treated with poly (I:C) or staurosporine the floating cells were centrifuged at 800 rpm for 5 min and combined later with the cells attached to the plate.

500 µl of TRI Reagent was added per well of the 6-well plate, pipetted up and down several times and cells from three wells were combined to one 2 ml centrifuge tube on ice. The tubes were spun at 13,000 rpm for 5 min at 4°C and the supernatant was transferred to a new 2 ml centrifuge tube. The samples were incubated for 5 min at room temperature. Then 300 µl of Chloroform (Fisher Scientific) was added and the tubes were vortexed for 20-30 s. After incubation at room temperature for 5-10 min and clear phase separation the samples were centrifuged at 13,000 rpm for 15 min at 4°C. After centrifugation 750 µl of the aqueous phase was carefully pipetted into a new 1.5 ml centrifuge tube without touching the organic phase. For isopropanol precipitation an equal volume of isopropanol was added to the aqueous phase. The samples were vortexed briefly and incubated at -20°C overnight.

The next day the samples were centrifuged at 13,000 rpm for 15 min at 4°C. In order to remove salt, the RNA pellet was washed twice with 80% ethanol and centrifuged each time at 13,000 rpm for 5 min at 4°C. After the last wash step the samples were spun again at 13,000 rpm for 1 min and the remaining drops were removed with 10 µl filtered tips. The RNA pellets were air-dried for 10 min, resuspended in 30 µl of RNase-free water and stored at -80°C.

The RNA concentration was measured using the Nanodrop 8000 spectrophotometer (Thermo Fisher Scientific). The quality of the RNA was deemed good if the 280/260 ratio was ~2.0, and the 260/230 ratio was between 2-2.2.

In order to check the quality and integrity of the RNA the samples were run on a 1.5% (w/v) agarose gel prepared in 0.5 x Tris-Borate-EDTA buffer (TBE) and stained with

10 mg/ml ethidium bromide (Fisher Scientific). For running the RNA on the gel 1 µg of RNA was added to 5 µl gel loading buffer II (Thermo Fisher Scientific, AM8546G). The buffer contains 95% formamide that stabilizes the RNA. Before running on the gel the RNA was denatured at 70°C for 2 min. The gel electrophoresis was performed at 100 V for 30 min and the gel was imaged using the Typhoon Scanner (GE Healthcare). The integrity of the RNA was assessed by checking the 28S and 18S ribosomal RNA bands.

2.4 Detecting small RNAs by Northern blotting

2.4.1 Separation of small RNAs by denaturing polyacrylamide gel electrophoresis

Before the preparation of the denaturing polyacrylamide gels all the equipment was soaked for at least 30 min with 10% Distel in order to destroy any RNases. The separation of small RNAs was performed using 15% denaturing polyacrylamide gels. To prepare a single gel 2.1 g urea (Fisher) was measured and 1.25 ml RNase-free water and 0.5 ml 5 x TBE was added to a 50 ml centrifuge tube. The mixture was heated in the microwave for 20-40s in order to dissolve the urea completely. After the solution cooled down, 1.85 ml 40% acrylamide/bis solution 19:1 (Biorad, 16110144) was added and the mixture was filled up with RNase-free water to a final volume of 5 ml. The centrifuge tube was gently inverted several times and the polymerisation reaction was initiated by adding 2.5 µl Tetramethylethylenediamine (TEMED) (Biorad, 1610800) and 50 µl of 10 % ammonium persulfate solution (APS) (Sigma, A3678) which acts as a radical initiator molecule in the polymerization process.

The gel solution was quickly poured between two 1.0 mm thin glass plates and a 10 well comb was inserted. The gel was polymerized for about 30 min. If more than 10 samples were loaded on the gel the gel solution was poured into a precast gel cassette and let it polymerize. The gel electrophoresis of small gels was performed using the mini PROTEAN III system (Biorad). The bigger gels with more than 10 wells were casted into Criterion cassettes (Biorad, 3459902) and the Criterion vertical electrophoresis cell (Biorad) was used for gel electrophoresis.

After the gel was polymerized the gel was pre-run in 0.5 x TBE at 100V for 30 min and the wells were washed with syringe and a 20-nauge needle.

For Northern blotting of small RNAs 10 µg of RNA was mixed with an equal volume of Gel loading buffer II (Thermo Fisher Scientific, AM8546G), denatured at 70°C for 2 min and kept on ice. Before loading on the gel, the wells were again rinsed with a syringe and a 20-nauge needle in order to remove some urea residues. The samples were carefully loaded on the gel and the gel electrophoresis was performed at 110 V for 2-3 h. After that, the gel was stained in a square petri dish with 10 mg/ml ethidium bromide in 0.5 x TBE for 5 min. The gel was imaged using the Typhoon scanner and RNA integrity and RNA separation was assessed before transfer.

2.4.2 Transfer of small RNAs onto a nylon membrane using semi-dry transfer

For the transfer six whatman filter paper (Thermo Fisher Scientific, 3030-335) and one piece of Hybond-NX nylon membrane (GE Healthcare, RPN202T) were cut at the same size as the gel and soaked in water. A neutral Hybond-NX membrane was used because positively charged nylon membranes are not compatible with the following chemical crosslinking using 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC).

For the semi-dry transfer the membrane and the gel were placed in between of six whatman filter paper. First three soaked whatman filter paper were positioned on the semi-dry apparatus (Fisher) and air bubbles were removed by a pipette. Then the membrane and the gel were placed on the three whatman filter paper followed by another layer of three whatman filter paper. Remaining air bubbles were carefully removed. The semi-dry transfer was performed at 20 V for 60 min at 4°C.

2.4.3 Chemical cross-linking

Before cross-linking the polyacrylamide gel was scanned using the Typhoon Scanner in order to ensure that all the RNA was transferred successfully. The small RNAs were crosslinked to the membrane by chemical cross-linking using EDC. It was previously shown that Carbodiimide-mediated cross-linking enhanced the sensitivity for the detection of small RNAs by Northern blotting up to 50-fold compared to conventional UV-cross-linking (Pall and Hamilton, 2008). For the cross-linking solution 122.5 µl 12.5 M 1-Methylimidazole (Sigma, M50834) and 10 µl concentrated hydrochloric acid (HCl) was added to 10 ml water. The pH of the solution was checked

using the pH meter and adjusted to pH 8.0. Then 0.373 g of EDC (Thermo Fisher Scientific, 22980) was added to the solution, gently mixed and filled up with water to a final volume of 12 ml.

A whatman filter paper that was slightly bigger than the membrane was soaked in cross-linking solution and placed on a Saran wrap. The membrane was placed with the RNA facing up on the soaked whatman filter paper, sealed tightly with Saran wrap and incubated in a hybridization oven for at 60°C for 1-2 h.

After cross-linking the membrane was washed twice in water on an orbital shaker for 5 min at room temperature, sealed with Saran wrap and stored at -20°C until hybridization.

2.4.4 Hybridization

For pre-hybridization the hybridization buffer (Sigma, H7033) was pre-warmed at 37°C. Then the membrane was placed in a hybridization tube containing 5-10 ml hybridization buffer while the RNA side of the membrane was facing inside the tube. The membrane was incubated at 37°C in a HB-1000 hybridization oven (UVP) under rotation.

While the membranes were incubated, the probe for Northern blotting was prepared in a screw-cap centrifugation tube. The antisense DNA oligonucleotide (without 5' end phosphate) was ordered and synthesized by Sigma. For preparing and end labelling the antisense DNA oligonucleotide probe 2 µl of a 10 µM oligonucleotide (Sigma), 2 µl of T4 Polynucleotide Kinase buffer (PNK) (NEB, M0201S), 14 µl nuclease-free water, 3 µl of [γ-^{32}P]-ATP (~1.1 MBq) (Perkin Elmer, BLU502A500UC) and 1 µl of T4 Polynucleotide Kinase (PNK) (NEB, M0201S) were gently mixed in a screw-cap centrifugation tube. PNK transfers the terminal phosphate from [γ-^{32}P]-ATP to the 5' end of the DNA oligonucleotide. The reaction was incubated in the radioactivity laboratory at 37°C for 1 h.

After 1 h 30 µl water was added to the tube and the whole reaction mixture was carefully added to the hybridization tube without touching the membrane. The membrane was incubated at 37 °C in the hybridization oven under rotation overnight.

2.4.5 Washing of the membrane and signal detection

Following overnight hybridization the membrane was washed twice with wash buffer containing 0.2 x Saline-Sodium citrate (SSC), 0.1% Sodium dodecyl sulphate (SDS) at 37°C for 30 min under rotation. If the background of the membrane was too high the washing was repeated in order to remove the non-specific radioactive signal. The membrane was wrapped in Saran wrap, placed in a cassette and exposed to a phosphorimager screen (Fujifilm) between 30 min and 5 days depending on the signal. The phosphorimager screen was imaged using the Typhoon Scanner and the Image Quant Software (GE Healthcare, Typhoon Scanner) was used for quantification.

2.4.6 Stripping of Northern blot membranes

In order to hybridize the membrane with another oligonucleotide probe, the hybridized labelled probe was removed from the membrane by incubating the membrane in 0.1% SDS for 30 min to 5 h at 90°C in a hybridization oven under rotation. Successful removal of the probe was checked by exposing the membrane to a phosphorimager screen and imaging using the Typhoon Scanner.

2.4.7 Oligonucleotide probe sequences used for Northern blotting

The following DNA oligonucleotides reverse complementary to the small RNA of interest (Sigma) were used for Northern blotting (Appendix II, **Table 0.12**). For some hY5 3' end mutants a locked nucleic acid (LNA) probe had to be used because the sequence was too short for the hybridization with a normal synthesized DNA oligonucleotide. In an LNA probe the LNA nucleotide (shown in Appendix II, **Table 0.12** by *) is modified by a methylene bridge between the 2' oxygen atom and the 4' carbon atom of the ribose. This modification locks the ribose in the 3' endo conformation by the bicyclic ring formation. The use of LNA probes increases the hybridization sensitivity and detection of small RNAs. The LNA probe was synthesized by Exiqon.

2.5 Western blotting

2.5.1 Polyacrylamide gel electrophoresis of proteins

For western blotting 10 µg of total protein extract was used. An equal volume of Laemmli buffer (Sigma, S3401) was added to the protein lysate, denatured at 100°C for 5 min and then immediately put back on ice until the gel was ready to be loaded. The proteins were separated on precast 4-12% gradient Bis-Tris polyacrylamide gels using the XCell Surelock mini-cell electrophoresis system (Invitrogen). The proteins were run in NuPAGE MOPS SDS running buffer (Thermo Fisher Scientific, NP0001) at 120 V for 1-2 h. As protein standard the PageRuler prestained protein ladder (Thermo Fisher Scientific, #SM0671) and MagicMark XP Western Protein Standard (Thermo Fisher Scientific, LC5603) were used.

2.5.2 Western blot transfer

After gel electrophoresis, the proteins were transferred onto a nitrocellulose membrane (Protran BA85, VWR, 732-4174) by wet-transfer using the XCell II blot module (Invitrogen).

For the transfer, six whatman paper were cut at the same size as the gel and soaked together with six sponge pads in 1x NuPAGE transfer buffer containing 10% methanol (Thermo Fisher Scientific, NP0006). For the wet-transfer first three soaked sponge pads and whatman filter paper were placed on the transfer machine, followed by the gel, nitrocellulose membrane and another layers of three whatman filter paper and sponge pads. The transfer was carried out at 300 mA for 2-4 h at 4°C.

2.5.3 Blocking of membrane and antibody hybridization

After the transfer the membrane was blocked for 1h at room temperature in TBS (Tris-buffered saline) buffer containing 0.1% Tween 20 (TBST) and 5% milk powder.

After blocking the membranes were incubated with the primary antibody in TBST-buffer containing 5% milk powder overnight at 4°C. The following day the membranes were washed three times in TBST buffer for 5 min at room temperature. After that, the membranes were incubated with a horseradish peroxidase (HRP) conjugated

secondary antibody in TBST-buffer supplemented with 5% milk powder for 1 h at room temperature. The membranes were washed again three times with TBST buffer and the western blot signal was detected using enhanced chemiluminescence (ECL) SuperSignal West Pico Chemiluminescence Substrate (Thermo Fisher Scientific, 34080). The Western blots were visualized using the SYNGENE gel documentation system.

2.5.4 Western blot antibodies

The antibodies used for the Western blots were diluted according to the manufacturer's instructions. The antibody for the detection of mouse Ro60 was kindly provided from the laboratory of Sandra Wolin and was described by Chen *et al*., 2003.

For the detection of β-actin the membrane was incubated with a primary antibody against mouse β-actin from Abcam (ab6276). In order to detect mouse Ro60 and β-actin a rabbit polyclonal secondary antibody against mouse IgG conjugated with HRP (Abcam, ab6728) was used.

2.6 Generation of DNA constructs

2.6.1 DNA extraction from HeLa cells

Genomic DNA was extracted from 2 x 10^6 HeLa cells using the GenElute Mammalian Genomic DNA Miniprep Kit (G1N10, Sigma) according to the manufacturer's instructions. The genomic DNA was stored in -20°C until use.

2.6.2 Amplification of the human Y5 RNA gene from genomic DNA

The human Y5 RNA gene was amplified from genomic DNA as described in (Turnbull, 2014).

As a template for polymerase chain reaction (PCR) 100 ng of genomic DNA was used. The PCR reaction was set up in a 20 µl reaction using Phusion High-Fidelity DNA polymerase according to the manufacturer's instructions (Thermo Fisher Scientific, F530L). For PCR set up 100 ng genomic DNA was mixed with a final concentration of 1 x Phusion HF (high fidelity) buffer, 0.4 mM dNTPs, 0.5 mM forward primer, 0.5 mM reverse primer (see Appendix II, **Table 0.1**) and 2 U Phusion DNA

Polymerase. In order to amplify the human hY5 gene including the RNA polymerase III promoter sequence and terminator region the forward primer was designed to bind specific to the promoter region and the reverse primer to bind to the terminator region of the human Y5 RNA gene. The human Y5 RNA gene was amplified including a 2.9 kb promoter region upstream the hY5 encoding region and 30 bp downstream of the terminator signal.

The PCR cycling conditions were as follows: 98°C for 3 min, followed by 30 cycles of 98°C for 1 min, 64°C for 45 s and 72°C for 3 min, followed by a final extension of 72°C for 10 min and 4 °C hold.

2.6.3 Agarose gel electrophoresis

The PCR products were separated by agarose gel electrophoresis. The percentage of the agarose gel varied between 0.8 % and 2% depending on the expected size of the PCR. The agarose was dissolved in 0.5 x TBE by heating up in the microwave. The gel was poured and 10 mg/ml ethidium bromide was added. The PCR products were mixed with 6 x DNA loading dye (NEB). The gel was run at 120 V for 30-60 min and imaged using the UV transilluminator or Typhoon Scanner.

2.6.4 Purification of PCR products by gel extraction

After visualization of the gel with the UV transilluminator the DNA bands were excised from the agarose gel with a sterile scalpel at the expected size and transferred to a 1.5 ml centrifugation tube. For the purification of PCR products from the agarose gel the Zymoclean Gel DNA Recovery Kit (Zymo Research, D4007) was used according to the manufacturer's instructions. The excised gel slices were dissolved at 55°C for 10 min and the DNA was eluted in 6-10 µl sterile water.

2.6.5 Adenosine tailing of PCR products for TA cloning

The High fidelity Phusion DNA polymerase generates blunt-ended DNA fragments. In order to use blunt ended DNA fragments for TA cloning, the PCR products need to have a 3' terminal deoxyadenosine. For the tailing reaction 1-7 µl purified PCR product was mixed with dATP at a final concentration of 0.2 mM, 1 µl of GoTaq G2 DNA Polymerase 10 x reaction buffer containing $MgCl_2$ and 5 U of GoTaq G2 DNA polymerase (Promega, M7801) in a final volume of 10 µl. The reaction mixture was incubated at 70°C for 30 min.

2.6.6 Cloning of PCR products into pGEMT easy vector

The pGEMT easy vector system I (Promega, A1360) was used for TA cloning of the purified PCR products. The pGEMTeasy vector (see Error! Reference source not found.) is a high-copy vector system that is linearized and contains 3' terminal deoxythymidine at both ends. The 3'-T overhangs at the insertion site create overhangs for the efficient ligation of PCR products that have a 3' terminal deoxyadenosine. The multiple cloning region of the pGEMT easy vector is flanked by the T7 and SP6 RNA polymerase promoter. Another feature of the pGEMT easy vector is the β-galactosidase gene in the multiple cloning region which allows the blue/white screening of recombinant bacteria. If an insert was cloned successfully in the multiple cloning region of the pGEMT easy vector the coding sequence of the β-galactosidase gene gets disrupted which results in white colonies whereas bacteria that only contain empty vector without any insert will appear blue. The ampicillin resistance gene in the pGEMT easy vector is used as another selection marker to screen for positive recombinant bacteria.

The A-tailed PCR product of the hY5 RNA gene including a 2.9 kb pol III promoter upstream the hY5 RNA gene (**Figure 2.1** B, shown in dark blue) and a 30 bp terminator region downstream (shown in **Figure 2.1** in dark red) of the hY5 RNA gene (**Figure 2.1 B** shown in dark green) was ligated into the pGEMT easy vector (**Figure 2.1 A**) and the pGEM-hY5-WT plasmid was generated (**Figure 2.1 B**).

A

B

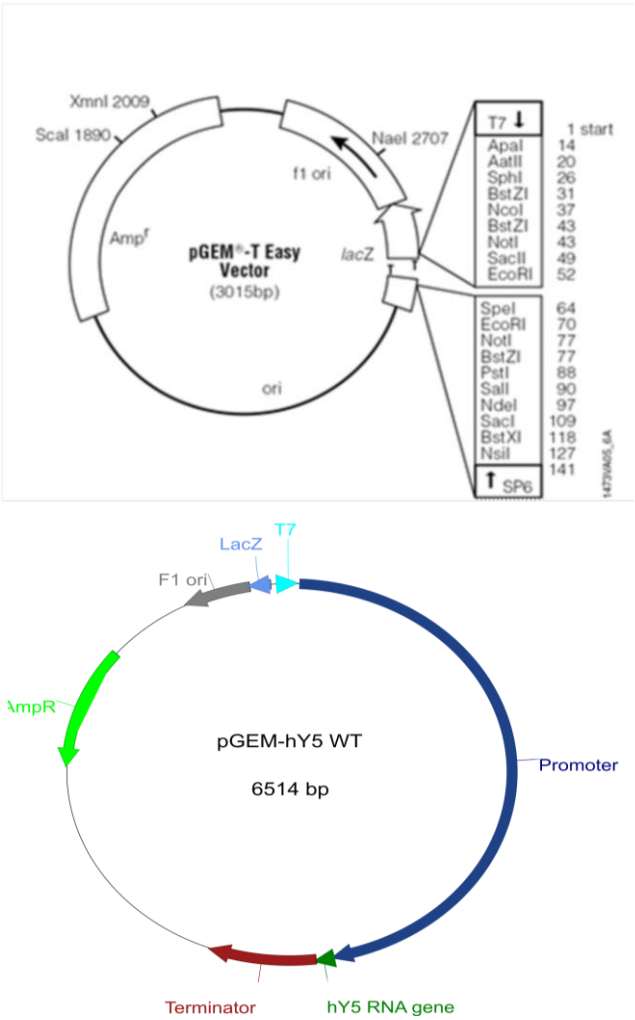


Figure 2.1. Vector map of pGEMTeasy and pGEM-hY5 WT plasmid DNA vector.

A. Vector map of pGEMT easy vector (Promega).

The vector map shows the circularized pGEMT easy vector with the origin of replication (ori), the Ampicillin resistance gene (Amp^r), the β-galactosidase gene (lacZ), the multiple cloning region including the restriction sites that is flanked by the T7 and SP6 promoter. The figure was taken from the pGEMT easy manual (Promega, A1360).

B. Vector map of the pGEM-hY5 WT plasmid.

The map shows the pGEM-hY5 plasmid that was generated by the ligation of the A-tailed PCR product of the hY5 RNA gene (shown in dark green) including a 2.9 kb RNA polymerase III promoter (shown in dark blue) upstream of the hY5 RNA gene (shown in green) and 30 bp terminator region downstream of the hY5 RNA gene (shown in red) into the pGEMT easy vector. The Amplicillin resistance (Amp) is shown in green. The LacZ gene for blue-white selection is shown in light blue. The T7 promoter is depicted in turquoise and the origin of replication is shown in grey. The vector map was generated using ApE the plasmid editor.

2.6.7 Preparation of supercompetent DH5α bacterial cells

An aliquot of DH5α *E.coli* bacterial cells was plated on a LB-agar plate and incubated at 37°C overnight. A single colony of DH5α cells was streaked onto a M9 minimum plate and incubated for 48-72 h. After that, a single colony was plated on an LB agar plate supplemented with 10mM $MgCl_2$. As a contamination control the single colony was also plated on an LB plate supplemented with 100 mg/ml Ampicillin. The plates were incubated at 37°C overnight. The next day the LB Ampicillin plate was checked for contamination and if nothing had grown a single colony from the LB/$MgCl_2$ plate was inoculated in 5 ml TYM media and incubated overnight at 37°C under shaking at 200 rpm. One ml of the overnight culture was added to 125 ml TYM broth in a 500 ml flask and incubated at 37°C under shaking at 300 rpm for an additional 3-4 h until the optical density (OD_{550}) reached 0.5. The culture was poured into 50 ml falcon centrifugation tubes and spun for 12 min at 1500 x g at 4°C.

The pellet was resuspended very gently in 40 ml of ice cold TfbI resuspending buffer and incubated on ice for 10 min. The suspension was centrifuged at 1500 x g for 8 min and resuspended in 4 ml of TfbII buffer. The supercompetent cells were aliquoted, frozen in liquid nitrogen and stored in -80°C. The competent cells were checked for transformation efficiency by transforming 0.1 ng of pUC19 vector control (NEB, N3041S). The transformation efficiency of DH5α cells was calculated to range between 1 x 10^7 to 1 x 10^8 cfu/µg of DNA.

2.6.8 Transformation of bacterial cells

For the transformation 100 µl of homemade supercompetent DH5α E. *coli* cells were thawed on ice for each transformation reaction. Then 5 µl of ligation reaction were added to each tube of competent cells, mixed gently and incubated on ice for 30 min. The mixture was then heat shocked at 42°C for 42 s and put on ice for 2 min. After that 900 µl of LB-media was added to the transformation reactions and incubated at 37°C with shaking for 1-2 h.

In order to determine positive clones with the insert by using blue-white screening, 0.5 mM X-Gal (5-bromo-4-chloro-3-indolyl-D-galactoside) and 80 mg/mL IPTG (Isopropyl-D-1-thiogalactopyranoside) were spread on the plates before the transformation reactions were plated. Of each transformation reaction 100, 150 and 200 µl were spread on LB plates supplemented with 100 mg/ml ampicillin. The bacteria plates were incubated at 37°C overnight.

2.6.9 Colony PCR

The transformation success and positive clones were determined by blue-white screening. Colony PCR was performed in order to confirm that the transformed bacteria contained an insert and if the insert is at the expected size. The pUC M13 forward and reverse primers were used for colony PCR. The pUC M13 forward and reverse primers bind to the flanking regions of the multiple cloning region and amplify the whole insert if it was successfully cloned into the pGEMT easy vector.

The PCR reaction was set up in a final volume of 10 µl using GoTaq G2 Flexi DNA Polymerase (Promega, M7801) according to the manufacturer's instructions.

The PCR reaction was set up with a final concentration of 1 x GoTaq Flexi DNA Polymerase buffer, 2 mM $MgCl_2$, 0.2 mM dNTPs, 0.2 mM forward primer, 0.2 mM reverse primer (see Appendix II, **Table 0.2**) and 0.5 U Phusion DNA Polymerase**.**

The PCR programme was chosen according to the manufacturer's instructions and the following conditions were used: 94°C for 2 min, followed by 29 cycles of 94°C for 30 s, 58°C for 30 s and 72°C for 90 s, followed by a final extension of 72°C for 10 min and 4°C hold.

The products of the colony PCR were analysed on a 1% agarose gel (110 V, 45 min) in order to confirm the presence and size of the insert. Positive colonies were inoculated in 5 ml LB-media supplemented with 100 mg/ml ampicillin as selection marker and incubated at 37°C overnight under shaking at 180 rpm.

2.6.10 Purification of plasmid DNA

The plasmid DNA purification was performed using the QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's instructions. The plasmid DNA was eluted in 30 µl water. The concentration of the plasmid DNA was measured using the nanodrop.

2.6.11 Sanger sequencing

In order to determine the sequence of the plasmids, the plasmid DNA sample was mixed at a concentration of 50-100 ng/µl with water to a final volume of 17 µl. For sequencing 2 µl of the sequencing primer (10 µM) was added and the samples were sent to Eurofins MWG Operon (Ebersberg, Germany) for Sanger Sequencing.

2.7 High throughput mutagenesis approach on the 3' end of the human Y5 RNA

2.7.1 Generation of the 3' end hY5 RNA mutant pool small RNA cDNA libraries

The generation of the 3' end hY5 RNA mutant pool cDNA libraries was performed by a previous PhD student (Turnbull, 2014). For the 3' hY5 mutagenesis analysis three different regions close to the 3' end cleavage site were chosen for random mutagenesis. In each of these regions five nucleotide substitution mutations were introduced resulting in 1024 possible combinations of hY5 mutants. The mutations were introduced using long primers containing random nucleotides, generating a pool of hY5 mutants for each of the three regions. The primers were synthesized in a way each of the bases was mixed equally with a proportion of 25% at each of the five mutated positions (IDT) in order to ensure that each hY5 mutant will be distributed equally. The PCR for mutant pool 1, 2 & 3 was performed in three replicates. The PCR products were A-tailed, ligated into pGEMT easy vector and transformed into E.coli. Statistical analysis revealed that for each mutant pool replicate a minimum of around 5000 colonies had to be harvested in order to be sure that each of the 1024 possible hY5 mutant will be represented at least once with a probability of 99% (Statistical analysis was performed by Dr. Irina Mohorianu and Carly Turnbull).

The colonies were harvested and the plasmid DNA was purified using the HiSpeed Midi Prep kit. The hY5 mutant pools 1, 2 and 3 were transfected into NIH 3T3 cells for 24 h. The cells were treated with staurosporine for 4 h. RNA was extracted and sRNA production was checked by Northern blot analysis.

The small RNA cDNA libraries were constructed using 2 µg of total RNA using High Definition (HD) adapters (Sorefan et al., 2012) with some modifications. For cDNA library construction a FAM-labelled 3' HD-adapter and 5' Illumina adapter were used. For PCR a longer RP-1 Index primer was designed in order to enrich for 3' end derived hY5 RNAs.

The sRNA cDNA libraries were sent for NGS to BaseClear (Netherlands) using a customized sequencing primer.

2.7.2 Generation of the 3' end hY5 RNA mutant pool full length cDNA libraries and plasmid libraries

Apart from the hY5 3' end mutant pools small RNA cDNA libraries, full length cDNA libraries and plasmid libraries were also prepared as controls (Turnbull, 2013). Full length hY5 RNA mutant cDNA libraries were generated in order to look at the transcription level and stability of hY5 RNA mutants. The libraries generated from the mutant plasmid pools itself were prepared to check if all the 1024 hY5 mutants are present and equally distributed.

The full length cDNA and plasmid libraries were sequenced on one lane using a customized sequencing primer. Since both full length and plasmid library sets have the same sequence it was impossible to sequence these libraries with the standard Illumina protocol because for successful clustering at the Illumina sequencing the first 15 nucleotides need to be different. For this reason a so called balancing DNA was generated by PCR (see Appendix II, **Table 0.4**) using the Phusion protocol in order to introduce heterogeneity and create a variety of different sequences. The balancing DNA contained a 20 nt stretch where random nucleotides were incorporated therefore sequence heterogeneity was ensured.

2.7.3 Bioinformatics analysis of hY5 3’ end mutant sRNA libraries

The bioinformatics analysis was carried out by Dr. Simon Moxon. As a first step, the Fastq files were converted to Fasta format and all reads containing ambiguous bases (“Ns”) were discarded. HD adapters containing four degenerate nucleotides were used for 3’ end ligation and for each individual library replicate a different index primer was used for library preparation.

Next, the first 8 nt of the 3’ HD adapter (5’-TGGAATTC-3’) were identified using perfect sequence match and trimmed off. The four degenerate nucleotides of the 3’ HD adapter at the ligation site of each sRNA read were trimmed. From the 5’ end of sequencing reads the Illumina adapter sequence was removed. The adapter trimming was performed using the UEA sRNA Workbench (Stocks et al., 2012). A Perl script was written in order to generate all possible 3’ end hY5 mutant sequences for each mutated region 1, 2 and 3 resulting in a total of 3072 possible hY5 mutant sequences. The reads were mapped full length and without allowing any mismatches to the 3072 possible hY5 mutant combinations using PatMaN (Prüfer et al., 2008). Each library replicate was ranked by abundance and for further analysis the mean rank across all the three replicates was taken.

However, PCR duplication is a huge problem during NGS library preparation and occurs through biases in the PCR amplification step. In this bioinformatics analysis the PCR duplicates were not removed. In order to remove PCR duplicates the degenerate nucleotides of the 3’ HD adapter could be used to identify the PCR duplicates. For that, any identical sequences with the same random nucleotides that are likely to be PCR duplicates would be removed and collapsed. Then, the adapter sequences would be trimmed off and aligned to all possible hY5 mutant combinations.

2.7.4 Bioinformatics analysis of the full-length cDNA and plasmid libraries

The bioinformatics analysis was carried out by Dr. Simon Moxon. For the analysis a Perl script was used to generate all possible 3' end hY5 mutant sequences for each mutated region 1, 2 and 3 resulting in a total of 3072 possible hY5 mutant sequences. The reads were aligned to all the possible hY5 RNA mutants without mismatch using PatMaN. The plasmid libraries contained extra 6 nucleotide TAACAC before the hY5 coding sequence in order distinguish the plasmid library reads from the full length hY5 mutant cDNA library reads. Each library replicated was sorted by abundance and the mean rank of all the three replicates was used for further analysis.

2.7.5 Secondary structure predictions and visualization

For secondary structure prediction of hY5 RNA mutants the programme RNAfold under default parameters was used (Hofacker, 2003). The visualization of the predicted hY5 RNA mutants was carried out using Varna (Darty et al., 2009).

2.7.6 Generation of individual hY5 3' end mutants and loop mutants

After bioinformatics analysis the top ten most abundant sequences of each mutant pool were cloned individually into the pGEMT easy vector.

The sequences of each oligonucleotide that was used for generating individual hY5 3' end mutants and deletion mutants can be found in appendix II.

The PCR was set up in a final volume of 20 µl using Phusion High-Fidelity DNA polymerase according to the manufacturer's instructions (Thermo Fisher Scientific, F530L) with some modifications. 1 ng of wild type hY5 plasmid (pGEM-WT) was set up with 1 x Phusion HF buffer, 0.4 mM dNTPs and 2 U Phusion DNA Polymerase without primers for 10 cycles at a lower annealing temperature. After 10 cycles 0.5 mM forward primer, 0.5 mM reverse primer (see Appendix II, **Table 0.3**) were added to the PCR mixture and the annealing temperature was raised to 69 °C for another 25 cycles. The PCR had to be modified that way because long reverse primers spanning the downstream region of the human Y5 RNA had to be used for the PCR reaction in order to amplify the whole hY5 mutant sequence.

The cycling conditions for the first PCR reaction were the following: 98°C for 30 s, followed by 9 cycles of 98°C for 10 s, 54°C for 30 s and 72°C for 2 min, followed by a final extension of 72°C for 10 min and 4°C hold.

After adding the primers the PCR reaction was run for another 25 cycles under following conditions: 98°C for 30 s, followed by 25 cycles of 98°C for 10 s, 69°C for 30 s and 72°C for 2 min, followed by a final extension of 72°C for 10 min and 4°C hold.

The PCR products were run on a 1% agarose gel and purified using the Zymoclean Gel DNA Recovery Kit according to the manufacturer's protocol. The purified PCR products were tailed and ligated into pGEMT easy vector as described before. The ligation products were transformed into E. *coli*. The plasmid DNA was purified using the QIAprep Spin Miniprep Kit according to the manufacturer's instructions. After checking the sequence and successful mutation by Sanger Sequencing a single colony was inoculated in 5 ml LB medium supplemented with 100 mg/ml ampicillin for 5-8 h at 37°C and shaking at 200 rpm. After that the starter culture was diluted 1/1000 in 100 ml LB medium containing 100 mg/ml ampicillin as a selection antibiotics. The culture was grown at 37 °C with shaking at 200 rpm overnight. The next day the culture was centrifuged at 4,000 x g at 4 °C and the plasmid DNA was purified using the HiSpeed Midi Prep kit (Qiagen, 12643). The plasmid DNA was eluted in 500 µl water and the concentration was determined by nanodrop.

2.7.7 Transfection of hY5 mutants into NIH 3T3 cells

The transfection of hY5 mutants was carried out as described in 2.1.4. For the transfection 20 µg of plasmid DNA was transfected using Fugene 6 for 24 h. For each transfection experiment a mock transfection control and empty vector control was included.

2.7.8 Treatment of NIH 3T3 cells with staurosporine

In order to induce apoptosis and hY5 derived small RNAs the cells were treated with staurosporine for 8 h at a final concentration of 1 µM.

2.7.9 RNA extraction and Northern blot analysis

The RNA was extracted as described in section 2.3 and the RNA integrity was checked on a 1.5 % agarose gel. The RNA concentration was measured using the Nanodrop.

For Northern blotting 10 µg of RNA was used and the Northern blot protocol was followed as previously detailed in section 2.4.

2.8 High throughput mutagenesis approach on the 5' end of the human Y5 RNA

2.8.1 Generation of 5' end hY5 mutant pools-first strategy

For the 5' hY5 mutagenesis approach a similar strategy like for the 3' end hY5 mutagenesis was designed. Three different regions near the 5' hY5 RNA cleavage site were selected for mutagenesis. In each of these regions five nucleotides were substituted by random nucleotides that would result in 1024 different hY5 mutants per mutated pool. The random mutations were introduced by overlap extension PCR using primers containing degenerate nucleotides. The primers had to be synthesized in a way that the five random nucleotides at each of the mutated regions were hand mixed at an equal proportion of 25% in order to ensure each base is equally distributed and less bias will be introduced by PCR. The PCR for mutant pool 4, 5 and 6 was carried out in three replicates. For PCR first 1 ng of wild type pGEM-WT was used as a template to generate the two fragments (long and short) for overlap extension PCR. The PCR was performed according to the manufacturer's instructions. 1 ng of pGEM-hY5 WT was mixed with a final concentration of 1 x Phusion HF buffer, 0.4 mM dNTPs, 0.5 mM forward primer, 0.5 mM reverse primer and 2 U Phusion DNA Polymerase (see Appendix II, **Table 0.5**).

The cycling conditions for the long fragment of mutant pool 4, 5 and 6 were as follows: 98°C for 30 s, followed by 30 cycles of 98°C for 10 s, 54°C for 30 s and 72°C for 2 min, followed by a final extension of 72°C for 10 min and 4°C hold.

For generating the short fragment for mutant pool 4, 5 and 6 following PCR cycling conditions were used: 98°C for 30 s, followed by 30 cycles of 98°C for 10 s, 52°C for 30 s and 72°C for 30 s, followed by a final extension of 72°C for 10 min and 4°C hold.

The PCR products of the long fragment and short fragment were purified using Zymoclean Gel DNA Recovery Kit according to the manufacturer's instructions. The purified PCR products were quantified using the nanodrop. For setting up the overlap extension PCR 100 ng of the long fragment (2.9 kb) and 5 ng of the short fragment was used. First the PCR reaction was set up without the addition of primers and final extension for 5 cycles in order that both fragments can anneal to each other. 100 ng of long fragment and 5 ng of short fragment were added to 1 x Phusion HF buffer, 0.4 mM dNTPs and 2 U Phusion DNA Polymerase.

The PCR conditions were as follows: 98°C for 30 s, followed by 4 cycles of 98°C for 10 s, 52°C for 30 s and 72°C for 2 min and 4°C hold.

After 5 cycles 0.5 mM forward primer and 0.5 mM reverse primer (see Appendix II, **Table 0.8**) were added to the PCR reaction. The annealing temperature was raised to 69 °C and the PCR was run for another 25 cycles. The PCR was run under the following conditions for another 25 cycles: 98°C for 30 s, followed by 25 cycles of 98°C for 10 s, 69°C for 30 s, 72°C for 2 min, followed by a final extension of 72°C for 10 min and 4°C hold.

The PCR for mutant pool 4, 5 & 6 was performed in three replicates. The PCR products were A-tailed, ligated into pGEMT easy vector and transformed into *E.coli*. In order to check successful mutagenesis 50 purified plasmid DNA samples of each mutant pool were sent for Sanger Sequencing.

For generating the large mutant pools around 9000-10000 colonies per mutant pool was harvested in order to ensure that each mutant will be present. Previous statistical analysis for the 3' end hY5 mutagenesis approach revealed that for each mutant pool replicate a minimum of around 5000 colonies had to harvested in order to be sure that each of the 1024 possible hY5 mutant will be represented at least once with a probability of 99% (Statistical analysis was performed by Dr. Irina Mohorianu and Carly Turnbull). The colonies were harvested and the plasmid DNA was purified using the HiSpeed Midi Prep kit.

2.8.2 Generation of hY5 5' end mutant pool plasmid libraries

For generating the hY5 5' end mutant pool plasmid libraries 5 ng of each mutant plasmid pool replicate was used for PCR. In order to introduce the Illumina adapter sequences and indices for Illumina sequencing two separate PCR reactions were carried out. The first PCR reaction was set up according to the Phusion protocol (see Appendix II, **Table 0.6**) using 5 ng of template. 98°C for 30 s, followed by 25 cycles of 98°C for 10 s, 55°C for 30 s, 72°C for 30 s, followed by a final extension of 72°C for 10 min.

The samples were separated on an 8 % native polyacrylamide gel. For preparing the 2 gels 10 ml water, 1.5 ml 5 x TBE, 3 ml of 40 % acrylamide/bis solution 19:1 were mixed and filled up with water to a final volume of 15 ml. For polymerisation 7.5 µl TEMED and 150 µl 10% APS were added and the gels were set for 30 min. The PCR reaction were mixed with 5 x Novex Hi-Density TBE Sample buffer (Thermo Fisher Scientific, LC6678) and run in 0.5 x TBE at 120 V for around 2 h. The gels were stained with SYBR Gold (Thermo Fisher Scientific, S11194) in 0.5 x TBE with a dilution of 1/10000 for 1 min at room temperature and visualized using the Typhoon Scanner. The DNA bands were purified from the gels using sterile razor blades and the gel slices were placed into a gel breaker centrifugation tube and centrifuge at 20,000 x g for 5 min. The gel pieces were eluted with 1 x NEB2 buffer overnight at 4°C under rotation. The next day the soaked gel pieces were centrifuged in a Costar Spin X column (Corning, CLS8160-96EA) at 2,600 rpm for 5 min at room temperature

in order to filter out the solid gel debris. The eluted DNA was concentrated by ethanol precipitation overnight at -20°C by adding 2.5-3 volumes of 100 % ethanol, 1/10 of the eluted volumes 3M sodium acetate, pH 5.2 and 15 mg/ml glycoblue coprecipitant (Thermo Fisher Scientific, AM9515). The next day the samples were centrifuged for 30 min at 20,000 x g at 4°C. The pellet was washed once with 80 % ethanol and centrifuged for another 5 min at 20,000 x g at 4°C. The pellet was air dried for 5-10 min at room temperature and resuspended in 12 µl water.

For introducing the Illumina index primer sequences to the first PCR product another PCR was set up using the Phusion protocol and Illumina index primer sequences (see Appendix II, **Table 0.6**). The first purified PCR product was diluted to 1 ng and used as a template.

The PCR conditions were as follows: 98°C for 30 s, followed by 25 cycles of 98°C for 10 s, 55°C for 30 s, 72°C for 30 s, followed by a final extension of 72°C for 10 min and 4°C hold.

The PCR product was separated on a native 8 % PAGE gel and purified as described above in detail. The plasmid libraries were resuspended in 12 µl nuclease-free water. After quantification and verification by Sanger Sequencing the plasmid libraries were sent for next generation sequencing and bioinformatics analysis was carried out.

2.8.3 Generation of the hY5 5' end mutant pools- second cloning strategy

Next generation sequencing and bioinformatics analysis revealed that the hY5 wild type sequence was overrepresented in the 5' end mutant plasmid pools. To overcome this problem another cloning strategy was performed. The previous cloning was based on a plasmid containing the wild type hY5 full length sequence. The mutations were introduced by overlap extension PCR and the fragments were purified by gel extraction. One possibility could be that small traces of the full length hY5 wild type sequence was carried over to the stage of overlap extension PCR and this wild type sequence got amplified more efficiently in the overlap extension PCR compared to the mutant hY5 sequences.

The cloning strategy was changed in a way that two different plasmids that contained a disrupted version of the hY5 full length sequence were used as templates instead of the plasmid encoding the full length hY5 RNA sequence. The PCR reactions for the truncated wild type hY5 sequences to generate the two template plasmid constructs were set up using the Phusion protocol (see Appendix II, **Table 0.7**). 1 ng of wild type hY5 plasmid and primer sets to create disrupted hY5 wild type sequences were used.

The PCR conditions for plasmid DNA template 1 were as follows: 98°C for 30 s, followed by 25 cycles of 98°C for 10 s, 52°C for 30 s, 72°C for 2 min, followed by a final extension of 72°C for 10 min and 4°C hold.

The PCR conditions for plasmid DNA template 2 were as follows: 98°C for 30 s, followed by 25 cycles of 98°C for 10 s, 52°C for 30 s, 72°C for 30 s, followed by a final extension of 72°C for 10 min and 4°C hold.

The PCR products were purified using the Zymoclean gel extraction kit, A-tailed, ligated into pGEMT easy vector and transformed into E. *coli*. The plasmid DNA constructs were sent for Sanger Sequencing and checked that the full length hY5 wild type sequence was not over-represented anymore.

The two plasmid DNA constructs with the confirmed truncated version of the hY5 full length wild type sequence were used as templates. For generating the PCR products for the mutant pools 4, 5 and 6 the degenerate nucleotides for each region were introduced by overlap extension PCR. The long fragment was amplified according to the Phusion protocol using 1 ng of template 1 (see Appendix II, **Table 0.7**). For the PCR of the short fragment 1 ng of template 2 was used.

The PCR for plasmid DNA template 1 was run at the following conditions: 98°C for 30 s, followed by 25 cycles of 98°C for 10 s, 52°C for 30 s, 72°C for 30 s, followed by a final extension of 72°C for 10 min and 4°C hold.

The PCR products were purified and quantified using the nanodrop. The final overlap extension PCR reaction was carried out as previously described (2.8.1). The two fragments were annealed at 55°C and after 5 cycles the forward and reverse primers were added to the PCR reaction for another 25 cycles.

The PCR conditions were as follows 98°C for 30 s, followed by 4 cycles of 98°C for 10 s, 52°C for 30 s, 72°C for 2 min without final extension and 4°C hold.

After 5 cycles the 0.5 mM forward primer and 0.5 mM reverse primers (see Appendix II, **Table 0.8**) were added to the PCR reaction. The annealing temperature was raised to 69°C and the PCR was run for another 25 cycles.

The overlap extension PCR conditions were: 98°C for 30 s, followed by 25 cycles of 98°C for 10 s, 69°C for 30 s, 72°C for 2 min, followed by a final extension of 72°C for 10 min and 4°C hold.

The PCR products were purified using the Zymoclean gel extraction kit, A-tailed, ligated into pGEMT easy vector and transformed into E. *coli*. The generation of the large plasmid mutant pools was carried out as previously described. For each mutant pool replicate 30 colonies were selected and plasmid DNA was sent for Sanger Sequencing in order to check successful mutagenesis.

2.8.4 Generation of full length cDNA and plasmid libraries for the hY5 5' end mutant pools

The plasmid libraries for the hY5 5' end mutant pools were prepared as previously described using 1 ng of template for PCR. The purified plasmid libraries were sent for next generation sequencing (Earlham Institute, Norwich Research Park).

For the full length cDNA libraries 5 µg of RNA was treated with Turbo DNase (Thermo Fisher Scientific, AM2238) for 30 min at 37°C to eliminate all the genomic DNA. The DNase treated RNA was purified using the RNA clean and concentrator kit according to the manufacturer's instructions. The RNA was eluted in 10 µl and the concentration was determined by nanodrop. The reverse transcription reaction was carried out using the Superscript II protocol according to the manufacturer's instructions. For first-strand synthesis 500 ng purified RNA was mixed with 1 µl of 10 mM dNTPs and 1 µl

of 2 µM hY5 RNA specific primer. The reaction was incubated at 65°C for 5 min and chilled on ice immediately. After denaturing the RNA, 4 µl of 5 x First Stand buffer, 2 µl of 0.1 M DTT and 1 µl of RNaseOUT (Thermo Fisher Scientific, 10777019) were added and incubated at 42°C for 2 min. At the final step 1 µl of Superscript II reverse transcriptase (Thermo Fisher Scientific, 18064014) was added and the reaction was gently mixed by pipetting up and down. The reaction was briefly centrifuged and incubated at 42 °C for 50 min. The reaction was inactivated at 70°C for 15 min. As described before for the plasmid libraries two PCR reactions had to be carried out in order to introduce the Illumina adapter and index sequences for the following next generation sequencing. The cDNA was used as a template for the first PCR reaction and the PCR was performed using the primers shown in Appendix II, **Table 0.9** according to the Phusion protocol for 15 cycles. The resulting PCR product was separated on a 6% PAGE gel and purified as described above. Then, 1 ng of the purified PCR product was used for the second PCR to introduce the Illumina index sequences according to the Phusion protocol for 10 cycles. The final cDNA full length libraries were run on a 6% PAGE and purified. The full length hY5 cDNA libraries were quantified using the Typhoon Image Quant software and after verification by Sanger sequencing the 5' end hY5 full length cDNA libraries were sent for next generation sequencing.

2.8.5 Small RNA library construction of hY5 5' mutants

The small RNA libraries were generated according to our published protocol (Xu et al., 2015) and standard Illumina protocol with some modifications. The total RNA was extracted using the TriReagent protocol as described previously. For small RNA library construction 2 µg of total RNA was used. The small RNA libraries were constructed with Illumina adapters and a customized RP1 PCR primer containing the original Illumina RP1 index sequence plus fifteen additional nucleotides on the 3' end of the oligonucleotide (see Appendix II, **Table 0.10**). This modified RP1 primer had to be used in order to enrich for YsRNAs because of their low abundance compared to other small RNAs. Different enrichment sequences for YsRNAs (5 nt, 10 nt and 15 nt) were tested and 15 nt additional hY5 RNA sequence seemed to be sufficient for YsRNA enrichment.

For small RNA library construction the 3' Illumina adapter (Sigma) had to be phosphorylated and adenylated because the truncated T4 RNA Ligase 2 (NEB) specifically ligates pre-adenylated 5´ ends of DNAs or RNAs to 3´ hydroxyl groups of

RNAs. For phosphorylation 6 µl of 100 µM 3' Illumina adapter was mixed with 10 µl of 10 x T4 polynucleotide buffer (NEB), 10 µl of 10 mM adenosine triphosphate (ATP) (NEB), 72 µl of water and 2 µl of 10 U/µl T4 polynucleotide kinase (NEB) and incubated at 37°C for 30 min. The phosphorylated product was concentrated by ethanol precipitation overnight at -20°C.

After centrifugation at 20,000 x g for 10 min at 4°C the pellet was washed with 250 µl 80% ethanol and centrifuged at 20,000 x g for 5 min. The pellet was air dried for 10 min and resuspended in 12 µl nuclease-free water.

For adenylation of the phosphorylated 3' adapters the 5' DNA adenylation kit (NEB) was used according to the manufacturer's manual. The reaction was set up by adding 4.5 µl of the phosphorylated 3' adapter to 4 µl of 10 x 5' DNA adenylation buffer (NEB), 4 µl of 1 mM ATP (NEB), 4 µl of Mth RNA ligase (NEB) and 23.5 µl nuclease-free water. The reaction mix was incubated at 65°C for 1 h and the enzyme was inactivated at 85°C for 5 min. The adenylation reaction was cleaned up by phenol: chloroform extraction (Sigma) and following ethanol precipitation at -20°C overnight. After centrifugation at 20,000 x g at 4°C for 15 min, the pellet was washed twice with 80% ethanol and air-dried for 10 min at room temperature. The pellet was resuspended in 12 µl nuclease-free water. The concentration of the 3' adenylated adapter was determined by nanodrop and adjusted to a final concentration of 20 µM. The successful adenylation was checked along with non-adenylated adapters on a 15% denaturing polyacrylamide gel electrophoresis using the mini PROTEAN III system and 1.0 mm gel plates (Bio-Rad).

The first step of the cDNA library construction was the ligation of the 3' adenylated adapter to the sRNA using truncated T4 RNA Ligase 2 (NEB). Therefore, 50% polyethylene glycol (PEG) 8000 solution (NEB) was warmed up at room temperature. After that, 8 µl of the RNA sample were mixed with 1 µl of 20 µM adenylated 3' adapter and denatured at 70°C for 2 min. The ligation reaction was set up by adding 2 µl of 10 x T4 RNA ligase 2 buffer, 0.75 µl RNaseOUT (Thermo Fisher Scientific), 4 µl 50% PEG solution (NEB), 1 µl T4 RNA ligase 2 and 3.25 µl nuclease-free water to the denatured RNA and adapter reaction mixture.

The ligation reaction was incubated for 2-3 h at 26°C and was cleaned up using the RNA clean and concentrator-5 kit (Zymo Research). The ligation product was eluted in 12.1 µl nuclease-free water and put on ice.

In order to further clean up and remove excessive 3' end Illumina adapter the ligation reaction was separated on an 18% denaturing PAGE along with appropriate size

markers and ligation controls. The gel was run at 100 V for 2-3 h, stained with SYBR Gold and imaged with the Typhoon Scanner. The ligation product was gel extracted and eluted in 1 x NEB2 buffer overnight as described before. The gel pieces were centrifuged in a Costar Spin X column (Corning, CLS8160-96EA) at 2,600 rpm for 5 min at room temperature and the DNA was concentrated by ethanol precipitation for 1-2 h in -80°C. The samples were centrifuged for 30 min at 20,000 x g at 4°C and the pellet was washed once with 80 % ethanol and centrifuged for another 5 min at 20,000 x g at 4°C. The pellet was air dried for 5–10 min at room temperature and resuspended in 3.5 µl water.

After the removal of excessive 3' Illumina adapter by gel purification the ligation product was ligated to the 5' end Illumina adapter which is a RNA adapter. The 5' end adapter was denatured at 70°C for 2 min. The 5' end ligation reaction was set up by adding 0.5 µl of 10 x T4 RNA ligase 1 buffer (NEB), 0.5 µl of 10 mM ATP (NEB), 0.5 µl of denatured 5' Illumina adapter and 0.5 µl T4 RNA ligase to the first ligation product. The ligation reaction was incubated for 2-3 h at 26°C.

After the ligation the di-tagged RNA was reverse transcribed using Superscript II (Thermo Fisher Scientific). The ligation reaction was mixed with 1 µl RTP RNA RT primer and denatured at 70°C for 2 min. Then 0.5 µl of 12.5 mM dNTPs, 1 µl of 0.1 M DTT, 1 µl of RNase Out and 1 µl of Superscript II was added to the reaction mixture and incubated at 50°C for 1 h.

For PCR amplification 4 µl of cDNA was mixed with 4 µl 5 x Phusion high fidelity buffer, 0.5 µl of 10 mM dNTPs, 1 µl of modified RP1 primer, 1 µl of barcoded RP index primer, 9.3 µl water and 0.2 µl of Phusion DNA polymerase. The PCR reactions were run at the following conditions: 98°C for 30 s, followed by 10-21 cycles of 98°C for 10 s, 60°C for 30 s, 72°C for 15 s, followed by a final extension of 72°C for 10 min and 4°C hold.

After PCR amplification with the optimal number of reaction cycles the PCR products were loaded on 8% non-denaturing PAGE gels along with a 20 bp size marker. Products between 140-150 bp were purified as described before.

2.8.6 Generation of individual hY5 5' end mutants

In order to validate the sequencing results after bioinformatics analysis several hY5 mutants were cloned individually. The PCR was set up as described in 2.8.3. The oligonucleotides used for the generation of individual hY5 5' end mutants can be found in the appendix II, Table 0.11.

The PCR products were tailed and separated on a 1% agarose gel. The PCR products were purified using the Zymoclean Gel DNA Recovery Kit according to the manufacturer's instructions. The PCR products were ligated into pGEMT easy vector as described before. The ligation products were transformed into E. *coli*. The plasmid DNA was purified using the HiSpeed Midi Prep kit (Qiagen, 12643) and eluted in 500 µl nuclease-free water. The concentration of the plasmid DNA was determined by nanodrop and sent for Sanger Sequencing (Eurofins MWG Operon, Ebersberg) to confirm the sequence and successful mutagenesis.

2.9 Small RNA library construction using High Definition adapters

The total RNA of wild type and Ro60 $^{-/-}$ mouse embryonic stem cells was extracted using the TriReagent method as previously described. Twenty µg of total RNA was used for purification using the mirVana kit according to the manufacturer's instructions.

For small RNA library construction 1 µg of purified total RNA was used. The adenylation of the 3' HD adapter was performed the same way as previously described for the 3' end Illumina adapter.

The small RNA cDNA libraries using HD adapter had to be modified and were constructed after the same protocol as described for the YsRNA cDNA libraries. The only difference was that for the wild type and Ro60 $^{-/-}$ mouse embryonic stem cell sRNA cDNA libraries the commercial RP1 index primer and different barcoded RP index primers were used.

The PCR products were separated by 8% non-denaturing PAGE gels along with a 20 bp size marker. Due to the fact of low abundant sRNAs in mouse embryonic stem cells and previous sequencing results both the miRNA fraction (140-150 nt) and sRNA fraction derived from longer RNAs (160 nt) were excised and purified as previously described.

Chapter 3 Establishing sequence and structural requirements for 3' end human Y5 RNA cleavage

3.1 Introduction

3.1.1 Cellular stress response

Cellular stress response is a vital process that allows cells to adapt and respond immediately to various types of physiological and environmental stress conditions. Such adverse stressors include UV irradiation, DNA damage, temperature shock, nutrient starvation, oxidative stress, toxins, drugs, viral infection or endoplasmic reticulum stress (Holcik and Sonenberg, 2005; Kültz, 2005; Pakos-Zebrucka et al., 2016).

In order to respond to these stress conditions immediately, protect cells from further damage and maintain homeostasis the cells had to evolve tightly regulated cellular stress response mechanisms. These stress response mechanisms involve the activation of molecular chaperones, DNA damage repair, autophagy, cell cycle control, or in severe cases cell death (Leung and Sharp, 2010). Importantly, the reprogramming of gene expression at transcriptional and translational level during stress allows the cells to respond rapidly to environmental changes in order to ensure cell survival and prevent damage (Spriggs et al., 2010; Vihervaara et al., 2018).

In the last decade it became increasingly clear that regulatory non coding RNAs are involved in cellular stress response. For instance, miRNAs as key regulators of gene expression have been shown to mediate stress response in animals and plants (Leung and Sharp, 2007; Sunkar et al., 2007).

Work in our laboratory has shown that miR-395 in *Arabidopsis thaliana* (*A. thaliana*) is upregulated under sulphate starvation and expressed in a tissue- and cell type-specific manner (Kawashima et al., 2009). In order to investigate if there are novel stress-induced miRNAs in animals, various cell types were exposed to different stress conditions, such as heat shock, cold shock, glucose starvation and mimicking viral infection and compared to an untreated control sample.

3.1.2 Previous work and background of the project

In previous work conducted in our laboratory the human breast cancer cell line MCF7 was treated with the immunostimulant poly (I:C) in order to identify novel miRNAs involved in viral infection. Therefore, small RNA cDNA libraries for 19-24 nt sized small RNAs from poly (I:C) treated and untreated MCF7 cells have been constructed and sent for next generation sequencing. Surprisingly, the sequencing results showed that only three miRNAs (miR-1246, miR-1975 and miR-1979) were differently expressed. Further investigation revealed that miR-1975 was derived from the 3' end of hY5 RNA, whereas miR-1979 was generated from the 3' end of the hY3 RNA (Nicolas et al., 2012). Meiri *et al*. confirmed the same finding one year later and the miRNAs were removed from miRBase (Meiri et al., 2010).

Previous experiments showed that during apoptosis human Y RNAs are cleaved in a caspase-dependent manner, generating Y RNA fragments with sizes ranging between 22-36 nt from the 3' and 5' end (Rutjes et al., 1999a).

Northern blot analyses in our laboratory demonstrated that during stress human Y RNA fragments were upregulated in several cancerous and non-cancerous cell lines compared to the untreated conditions. Also it was noted that human YsRNAs were also generated under normal conditions at a detectable level. In order to investigate the biogenesis of YsRNAs it was shown in further experiments that YsRNA generation is Dicer independent and Y RNA fragments are not in a complex with Argonaute 2. Therefore, YsRNA generation seems to be independent of the canonical miRNA pathway (Nicolas et al., 2012). However, it is still obscure how these YsRNAs are produced at specific cleavage sites.

The sequence and secondary structure of Y RNAs are evolutionarily conserved (Mosig et al., 2007; Perreault et al., 2007). One interesting and important question is whether any specific sequence motif or secondary structure or both are required for the specific Y RNA cleavages to occur. In order to answer this question, the human Y5 RNA was chosen for the experiments, because the expression of hY5 RNA can be studied in mouse cells which only express endogenous mY1 and mY3.

A previous PhD student cloned the full length hY5 RNA including a 2.9 kb RNA polymerase III promoter upstream and a 30 bp terminator region downstream of the hY5 RNA into pGEMT easy vector and started to introduce mutations close to the Ro60 binding site. In this previously conducted work it was shown by deletion and substitution mutagenesis analyses on the conserved stem-loop structure of hY5 showed that Y RNA cleavage seemed to correlate with the secondary structure. It was demonstrated that the stem-loop structure and Ro60 binding site were required for the generation of YsRNAs (Turnbull, 2014).

Due to the limit of targeted mutagenesis analysis, a high throughput mutagenesis approach was designed by Turnbull (Turnbull, 2014) in order to be able to examine thousands of hY5 mutants within this particular region at the same time.

Three different regions (designated as library 1, 2 and 3) near the 3' end cleavage site of the hY5 gene were selected for random mutagenesis (**Figure 3.1 A and B**).

The three different plasmid mutant pools were transfected into 3T3 mouse cells and the cells were treated with staurosporine in order to induce apoptosis and YsRNA cleavage products. Small RNA libraries from the resulting YsRNAs were constructed and sequenced through NGS (Figure 3.1) (Turnbull, 2014).

Preliminary bioinformatics analysis resulted in a list of fifty most expressed YsRNA cleavage products from three mutant pools. Interestingly, the NGS results indicated that the Y RNA cleavage site correlated to the secondary structure of the mutated hY5 RNAs. It was observed that the Y RNA cleavage of the mutated hY5 RNAs from library 2 and 3 showed wild type sized YsRNAs, while the Y RNA cleavage site was 2 nt upshifted in some mutants of library 1 compared to the wild type cleavage site. Concluding the results from the preliminary NGS analysis, it was hypothesized that the position of cleavage site is influenced by the secondary structure of the mutant hY5 RNA sequences (Turnbull, 2014).

However, in this previous work the sequencing results were not confirmed experimentally and the bioinformatics analysis had to be repeated because the data were not normalized correctly.

As part of my thesis I continued the project at this point in order to validate and further investigate sequence and structural requirements for human Y5 RNA cleavage. Therefore, the work undertaken in this chapter is a continuation of the 3' end hY5 RNA high throughput mutagenesis experiments which were performed by Carly Turnbull (Turnbull, 2014).

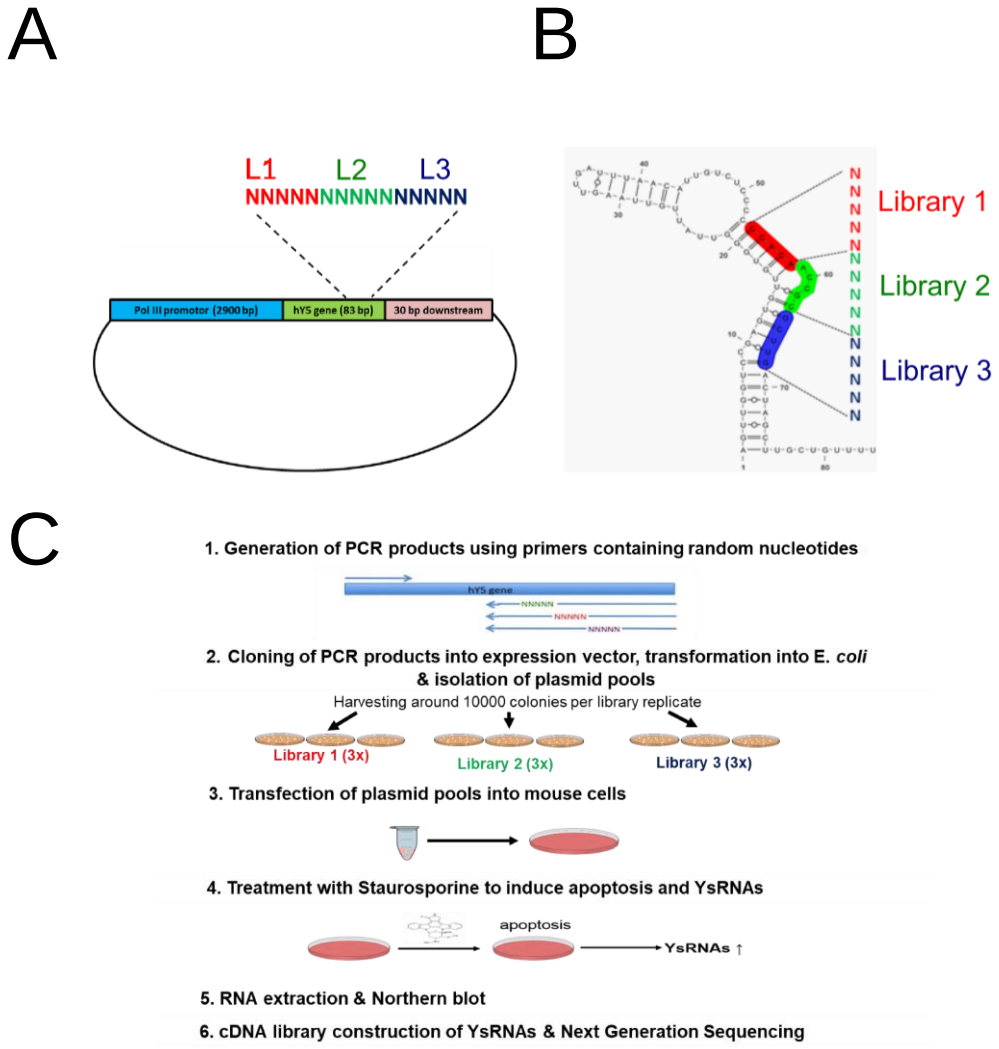


Figure 3.1. Overview of the design and workflow of the 3' end high throughput mutagenesis on the 3' end of the hY5 RNA.

A. Schematic of the pGEM-hY5-WT vector containing the mutated regions of hY5 RNA library 1, 2 and 3.

The mutations for library 1 (red), library 2 (green) and library 3 (blue) were introduced by mutagenesis PCR using primers containing random nucleotides (N).

B. Predicted secondary structure of human Y5 RNA including the regions that were selected for random mutagenesis at the 3' end.

In each of these regions of library 1 (red), library 2 (green) and library 3 (blue) five nucleotide substitution mutations were introduced resulting in 1024 (4^5) possible combination for each library.

C. Workflow of the 3' hY5 RNA mutagenesis approach.

The three different plasmid mutant pools were transfected into 3T3 mouse cells and the cells were treated with staurosporine in order to induce apoptosis and YsRNA cleavage products. Small RNA libraries from the resulting YsRNAs were constructed and sequenced through NGS.

In this chapter, I mainly performed the validation of the next generation sequencing results from the 3' hY5 mutagenesis analysis. From bioinformatics analysis which was performed by Simon Moxon I obtained a list containing the most abundant and least abundant hY5 RNA mutants of mutant pool library 1, 2 and 3. In this thesis, I did a further in-depth analysis of the resulting secondary structures of the thirty most and least abundant hY5 RNA mutants. From this analysis I generated and cloned the ten most and least abundant hY5 RNA mutants of each library into an expression vector and tested their expression levels and YsRNA production during stress by Northern blot analysis.

In order to investigate further sequence and structural elements that are involved in hY5 RNA cleavage from the 3' end of hY5 RNA I additionally performed mutagenesis on the internal loop L2a. My aim was to explore if the length or the nucleotide sequence of this loop has an influence on hY5 RNA cleavage from the 3' end.

As a result, I could validate and establish novel sequence and structural requirements for human Y5 RNA cleavage from the 3' end.

3.2 Results

3.2.1 Bioinformatics analysis of 3' hY5 mutagenesis libraries

All the cDNA libraries for the 3' hY5 high throughput mutagenesis were generated in three replicates and sent for NGS (Turnbull, 2014). The libraries were sequenced on the Illumina HiSeq 2000 at BaseClear (Netherlands). The bioinformatics analysis was carried out by Dr. Simon Moxon.

In this section I am giving an overview of all the analyzed datasets for the 3' end hY5 RNA mutagenesis study:

The sRNA cDNA libraries of mutant hY5 RNAs were prepared in order to determine all YsRNAs that were derived from full length hY5 RNA sequences during apoptosis. Apart from the sRNA libraries, libraries from mutant plasmid pools and cDNA of full-length hY5 RNA sequences were prepared for the normalization of the sRNA data set. The purpose of creating libraries from each plasmid pool was to check if all the hY5 RNA mutant sequences are present and to determine if some hY5 RNA mutant sequences were cloned more efficiently than others. The libraries from the full-length hY5 RNA sequences were created to see if all mutant hY5 RNA sequences were transcribed in the cells.

The analyses of the plasmid libraries and the full-length hY5 libraries allowed us to determine if the random mutagenesis was successful. With the analysis of these data sets we checked if the mutagenesis was random and if there was a strong selection for a particular hY5 RNA mutant sequence or if there was a problem that not all mutant hY5 RNA sequences were successfully transfected.

Further, with the bioinformatics analysis of the plasmid bias and full length hY5 RNA sequencing results it was possible for us to determine if there is any common feature among all the hY5 RNA mutants that were successfully transfected and transcribed. Therefore, the full-length hY5 RNA libraries and plasmid libraries were used to look at biases introduced by PCR, transfection and transcription.

3.2.1.1 Analysis of plasmid libraries

The plasmid libraries for the 3' end high throughput mutagenesis approach were generated in order to check if the random mutagenesis was successful or a particular hY5 RNA mutant sequences was over- or underrepresented. The plasmid libraries

were performed in three biological replicates. By introducing Illumina adapter sequences on the 3' and 5' end of hY5 RNA plasmid sequences the plasmid libraries could be sequenced using the Illumina NGS platform.

For the analysis of the plasmid libraries the sequencing reads were mapped to all possible 1024 hY5 mutant full length hY5 RNA sequences using PatMaN (Prüfer et al., 2008) without allowing any mismatch and a custom Perl script was used to parse the mapping file and count the number of unique motifs obtained. The obtained sequencing reads are summarized in **Table 3.1**.

The bioinformatics analysis of the plasmid libraries showed that when all the reads were mapped to all possible 1024 (4^5) full length hY5 RNA sequences that for each library replicate we got all expected 1024 hY5 RNA mutant sequences and the abundance of each sequence motif was more or less equally distributed. From the bioinformatics analysis we could see that there was no particular hY5 RNA mutant sequence that was overrepresented. The results from the bioinformatics analysis confirmed that the random mutagenesis approach that was performed on the 3' end of hY5 RNA was successful. All 1024 different hY5 Y RNA mutant sequences were cloned successfully and there was no significant bias against any particular sequence motif.

3.2.1.2 Full length cDNA libraries

The full length cDNA libraries were prepared in order to investigate if all the hY5 RNA mutants were successfully transfected and transcribed or if some hY5 RNA mutant sequences were missing. The full length cDNA libraries were generated in three biological replicates. The Illumina adapter sequences were introduced by PCR. Using different Illumina barcode index sequences the full length cDNA libraries were sequenced on the Illumina HiSeq 2500 sequencing platform. After adapter removal the obtained sequencing reads of the full-length cDNA libraries were mapped to all possible 1024 hY5 mutant full length hY5 RNA sequences using PatMaN without allowing any mismatch and again analysed using a custom Perl script. The obtained sequencing reads are summarized in **Table 3.2.**

Table 3.1. Overview of sequenced 3' end hY5 plasmid bias libraries.

The table summarizes the redundant read numbers for biological replicate of plasmid bias (PB) library 1, 2 and 3 and unique sequencing reads mapping to each possible full length hY5 RNA mutant sequence. The first column shows the three biological replicates of each plasmid bias library 1, 2 and 3 that were generated and sent for NGS. The second column shows the number of total sequencing reads for each plasmid bias library replicate after mapping the sequencing reads to all possible 1024 hY5 mutant full length sequences.

	Total reads	Unique hY5 RNA sequence motifs
PB 1-1	592365	1024
PB 1-2	572583	1024
PB 1-3	480579	1024
PB 2-1	482195	1024
PB 2-2	425736	1024
PB 2-3	488713	1024
PB 3-1	304449	1024
PB 3-2	336302	1024
PB 3-3	698035	1024

Table 3.2. Sequencing information of 3' end hY5 full length cDNA libraries.

The table shows the redundant read numbers for biological replicate of full length (FL) cDNA library 1, 2 and 3 and unique sequencing reads mapping to each possible full length hY5 RNA mutant sequence. The first column shows the three biological replicates of each full length library 1, 2 and 3 that were generated and sent for NGS. In the second column the number of total sequencing reads for each full length library replicate after mapping the total reads to all possible 1024 hY5 mutant full length sequences is listed. In the full length cDNA library 3 replicates not all 1024 unique mutant sequences were present and the number of mutant sequences missing is shown in blue.

	Total reads	Sequence motifs
FL 1-1	1414197	1024
FL 1-2	1388192	1024
FL 1-3	1780207	1024
FL 2-1	1284671	1024
FL 2-2	1023731	1024
FL 2-3	748520	1024
FL 3-1	1099655	984 (40 missing)
FL 3-2	775323	906 (118 missing)
FL 3-3	872712	978 (46 missing)

The bioinformatics analysis of the libraries for mutant pools 1 and 2 showed that all the hY5 RNA mutant sequences were transcribed successfully and that the generated transcripts were stable.

Intriguingly, in the full-length cDNA libraries of mutant pool 3 hY5 RNAs many motifs were not represented across the three replicates. In replicate FL 3-1 40 sequence motifs were missing whereas in FL 3-2 and FL 3-3 118 and 46 sequence motifs were absent. It has to be noted, that the mutated region of mutant pool 3 was very close to the Ro60 binding site (**Figure 3.2**, **A**, shown in blue).

We know from the literature that the bulged helix of the Ro60 binding site is highly conserved and Ro60 was shown to interact with the major groove of Y RNAs. In ribonuclease protection experiments on isolated RoRNPs from Hela cells it was demonstrated that the Ro60 protein binding site contains a single bulged cytidine within a helix (Wolin and Steitz, 1983). Further chemical probing experiments and mutagenesis analysis of hY1 RNA revealed that a bulged cytidine residue is important for Ro60 binding and interaction. Deletion of the cytosine bulge resulted in the inhibition of Ro60 binding. Interestingly, the mutation of the cytosine bulge to adenosine still allowed Ro60 binding but around five times less efficiently (Pruijn et al., 1991).

A few years later Green *et al*. investigated the features of Y RNA sequence and structure that are necessary for protein recognition and they found that specific base pairs in the conserved bulged helix are critical for Ro60 protein recognition. Chemical probing and mutagenesis experiments of xY3 RNA in *X. laevis* demonstrated that a single bulged nucleotide and a bulge of 3 nt on the opposite strand of the helix are essential for Ro60 binding. Ro60 was shown to bind to the lower stem region of the conserved stem of Y RNAs and base pairing between C9 and G90 as well as a bulge at position 8 was essential for Ro60 recognition (Green et al., 1998).

After the observation that some of the library 3 mutant motifs were absent, I performed sequence logo analysis of the missing motifs and compared it to the sequence logos of the fifty most abundant library 3 sequence motifs (Figure 3.2, B). For sequence logo analysis all the three replicates were combined and the programme WebLogo Version 2.8.2 (Crooks et al., 2004) was used for sequence logo generation.

From the sequence logo analysis I could observe that the fifth position of the pool 3 mutated region shows an exclusive preference for guanosine (**Figure 3.2**, **B**). This suggested that a G-C base pair between the positions 8 and 68 is necessary in order to retain the cytidine bulge and ensure Ro60 binding as it was already reported in the literature.

In position four of the sequence logo uridine seems to be preferred compared to guanosine, adenine and cytidine. An explanation for this G-U wobble base pair could be that this position requires some flexibility in order that a single nucleotide bulge in the major groove of the helix can be formed and Ro60 binding can occur.

The secondary structure analysis of the missing sequences revealed that the mutant pool 3 sequence motifs that were missing in the libraries change the secondary structure of the mutant hY5 RNA sequence in a way that they lack a single cytidine bulge and the bulge increased to an internal loop (**Figure 3.2**, **D**). Importantly, in those mutants the G-C base pair at position 8 was absent. Therefore, the Ro60 binding site in all missing full length hY5 mutant sequence motifs was disrupted and the stability of the hY5 mutant RNA was clearly affected (**Figure 3.2**, **D**) in comparison to a highly abundant pool 3 mutant (**Figure 3.2**, **C**).

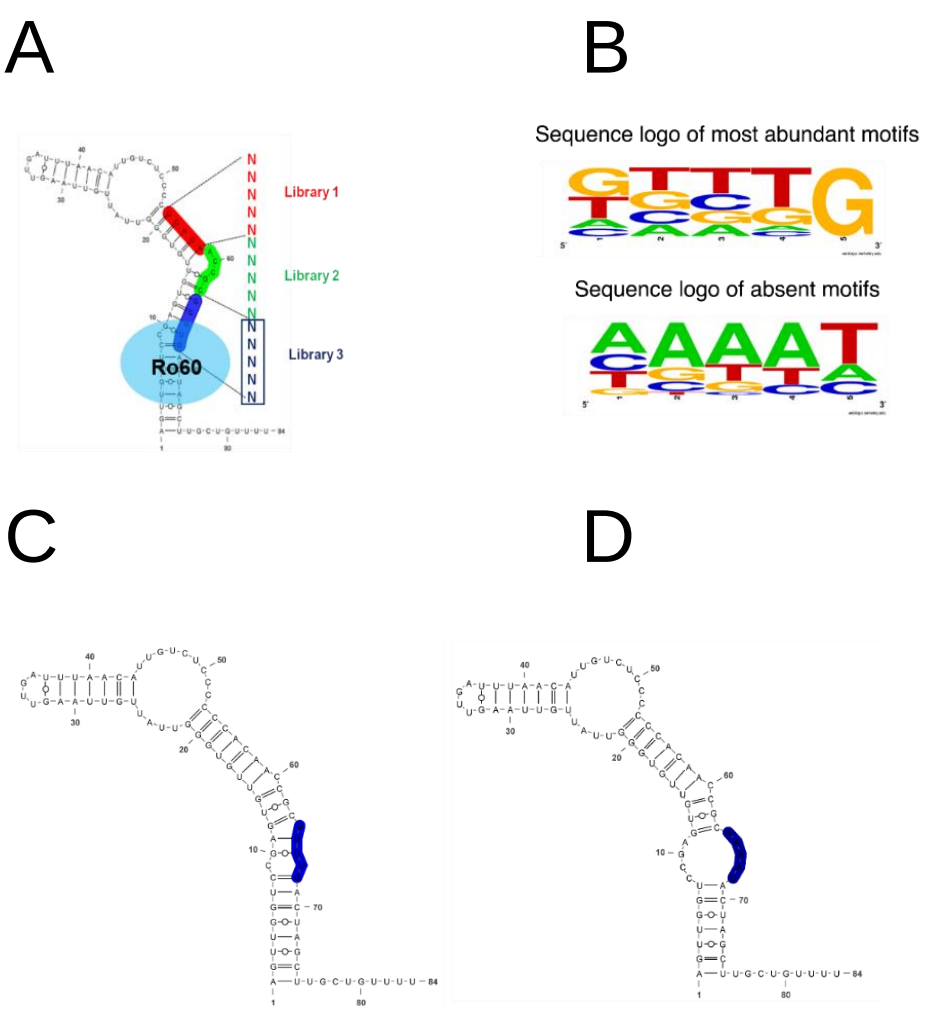


Figure 3.2. Requirement of a single cytosine bulge and a G-C base pair between the positions 8 and 68 for Ro60 binding and hY5 RNA stability.

A. Predicted secondary structure of human Y5 RNA.

The mutated region of library 3 is shown in dark blue. The Ro60 binding site is indicated in light blue.

B. Sequence logo analysis of most abundant and absent hY5 mutant pool 3 sequence motifs.

The position 5 of abundant mutant pool 3 sequence motifs has a strong G preference.

C. Predicted secondary structure of a highly abundant mutant pool 3 sequence motif.

The mutated region is shown in blue. In this highly abundant pool 3 mutant the secondary structure is similar to the wild type hY5 RNA with an intact cytidine bulge and Ro60 binding site.

D. Predicted secondary structure of a highly abundant mutant pool 3 sequence motif.

The mutated region is shown in blue. The sequence motifs missing in the full length mutant pool 3 libraries result in a change of secondary structure. The cytidine bulge and G-C base pair are not present anymore which affects the Ro60 binding site and stability of missing mutant pool 3 full length hY5 RNAs.

3.2.1.3 Small RNA cDNA libraries

During apoptosis full length Y RNAs get cleaved at the 3' end and 5' end of Y RNAs resulting in 3' end and 5' end derived YsRNAs. The YsRNA cDNA libraries were generated in order to investigate which 3' end hY5 RNA derived Y RNA fragments are more abundant than others and to examine which hY5 RNA full length mutant sequences are cleaved more or less efficiently due to different sequence motifs. The YsRNA cDNA libraries were prepared in three biological replicates. On the 3' end and 5' end of the YsRNAs two adapter sequences were ligated and multiplexing was performed using different Illumina barcode sequences. The libraries were sequenced on an Illumina Hiseq2500 platform.

The first step in the bioinformatics analysis of the YsRNA cDNA libraries involved the removal of the adapter sequence from the 3' end of the sequencing read. Next, the sequencing reads were mapped to the mouse reference genome. The reads mapping to the mouse genome were discarded because we were just interested in the reads mapping to the hY5 mutant sequences. In order to be able to map the obtained reads to the hY5 mutant sequences for each library a reference set containing all possible hY5 mutant sequences was generated using a Perl script. PatMaN was used to map the remaining reads of each library separately to the hY5 mutant sequences while no mismatch was allowed.

For library 1 sRNA libraries we obtained between 2 and 3 million reads, for library 2 libraries between 2.7 and 6 million, whereas for library 3 libraries we got between 1.1 and 3.8 million total reads. After adapter removal between 57% in and 75% of the reads mapped to the mouse genome (GRCm38/mm10, USCS genome browser). From the remaining reads between 2 % and 13% mapped to the hY5 mutant sequences, whereas for library 3 sRNA library most of the sequencing reads mapped to hY5 RNA (shown in Appendix III, **Table 0.13**).

From the size class distribution of the YsRNA libraries (**Figure 3.3***,* **A**) I could see that the size of YsRNAs derived from hY5 library mutant pool 2 and 3 (shown in green and blue) is mainly 32-33 nt whereas hY5 mutant pool 1 Y RNAs (shown in red) generate longer YsRNAs with a length of 34-35 nt. The start position of the sequencing reads of hY5 mutant pool 2 and 3 indicates that the cleavage occurs between T49 and C50 or C50 and C51. This finding is consistent with previous sequencing results obtained from MCF7 cells treated with poly (I:C) in which most of the 3' end derived hY5 RNA cleavage products were 32 nt.

3.2.2 Some library 1 hY5 RNA mutants generate longer YsRNAs compared to hY5 mutant pool 2 and 3

For all YsRNA reads from each library replicate a size class distribution was generated. Interestingly, the size class distribution of all YsRNA sequencing reads revealed that some library 1 mutant YsRNA reads were 1-2 nt longer compared to hY5 mutant pool 2 and 3 YsRNA reads (**Figure 3.3 A,** shown in red). Mutant pool 1 YsRNA reads were mainly 34/35 nt whereas mutant pool 2 and 3 YsRNAs were mainly between 32/33 nts.

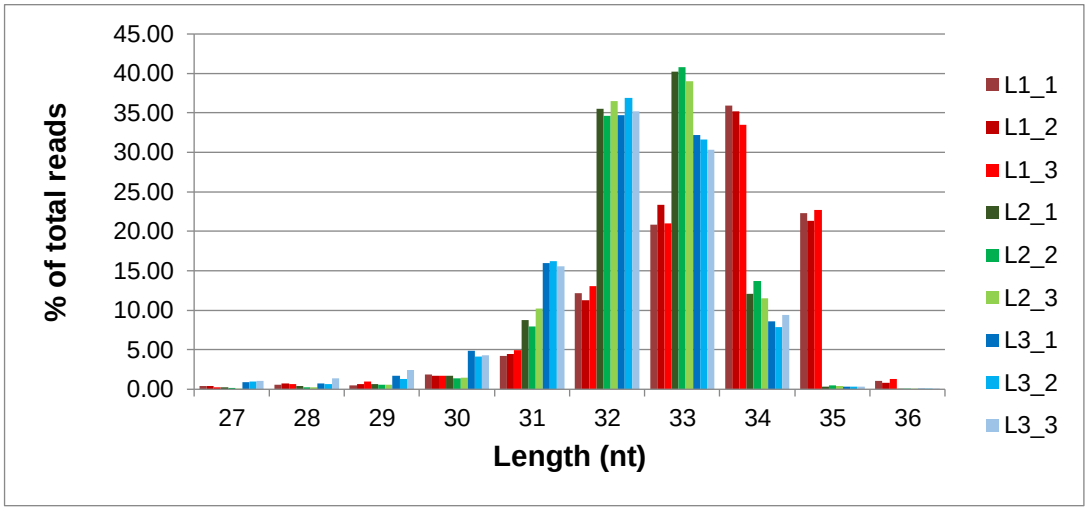
In order to investigate and validate these sequencing results the YsRNA reads were ranked by abundance. Then I looked at the respective predicted hY5 mutant RNA secondary structures if the introduced mutations affected RNA folding. I performed the RNA secondary structure analysis with RNAfold (Gruber et al., 2008) and for comparison I analyzed the 30 most abundant and 30 least abundant YsRNAs derived from each hY5 RNA mutant library pool.

From the RNA secondary structure analysis (**Figure 3.3**, **F and G**) I could show that the overall structure of the library 1 hY5 RNA mutants with the most abundant YsRNAs is similar to the wild type hY5 RNA (**Figure 3.3**, **D, E**).

However, the mutations introduced in the library 1 mutants that produced longer fragments changed the secondary structure in a way that the internal loop gets shorter and the Y RNA cleavage occurs 2-3 nt more upstream compared to the wild type hY5 RNA, but still two nucleotides above the conserved stem (**Figure 3.3**, **F**). Intriguingly, all the most abundant hY5 library 1 mutants have a stem S2 region (**Figure 3.3**, **D**) that consists at least of five Watson-Crick base pairs.

Interestingly, for the library 1 hY5 RNA mutant sequences with the least abundant YsRNAs it was shown from bioinformatics analysis that these mutants did not generate any YsRNAs (**Figure 3.3**, **B**). The secondary structure analysis of these mutants (**Figure 3.3**, **G**) revealed that their secondary structures folded in a different way than the wild type hY5 RNA.

A



B

Five most abundant motifs

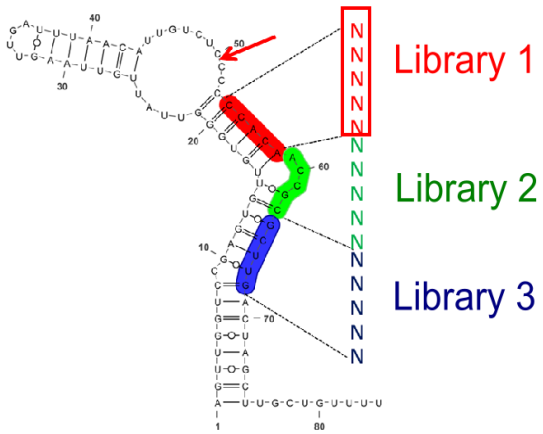
Mutant	Abundance
L1-1 ACGUC	3430
L1-2 ACCGC	2722
L1-3 CACAC	1852
L1-4 AUGGC	1743
L1-5 ACUGC	1529

C

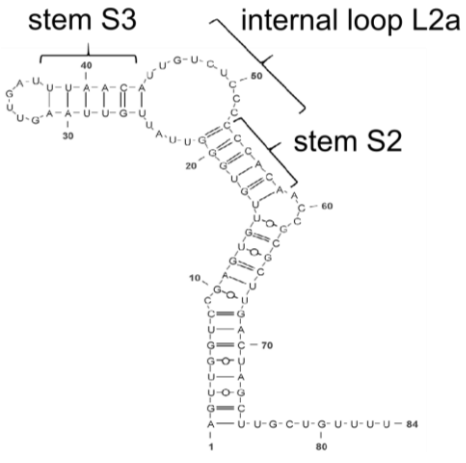
Five least abundant motifs

Mutant	Abundance
L1-6 GUUGC	21
L1-7 UCAAA	7
L1-8 UUGAA	5
L1-9 UCGCG	2
L1-10 UCAAG	1

D

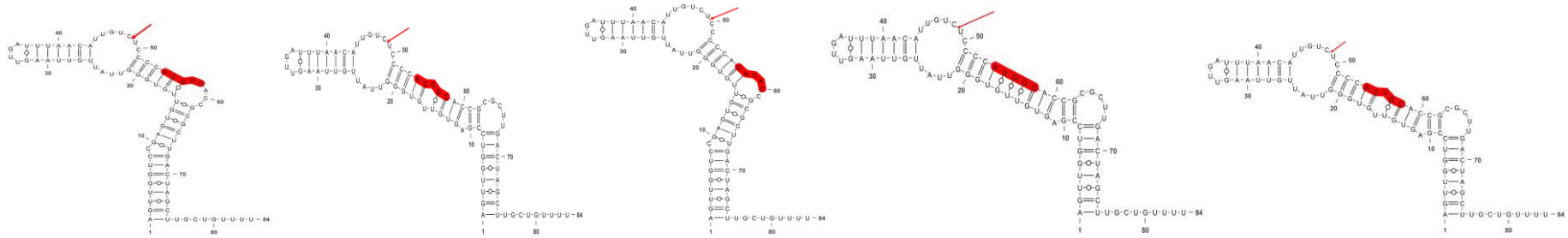


E



F

L1-1 ACGUC L1-2 ACCGC L1-3 CACAC L1-4 AUGGC L1-5 ACUGC



G

L1-6 GUCGC L1-7 UCAAA L1-8 UUGAA L1-9 UCGCG L1-10 UCAAG

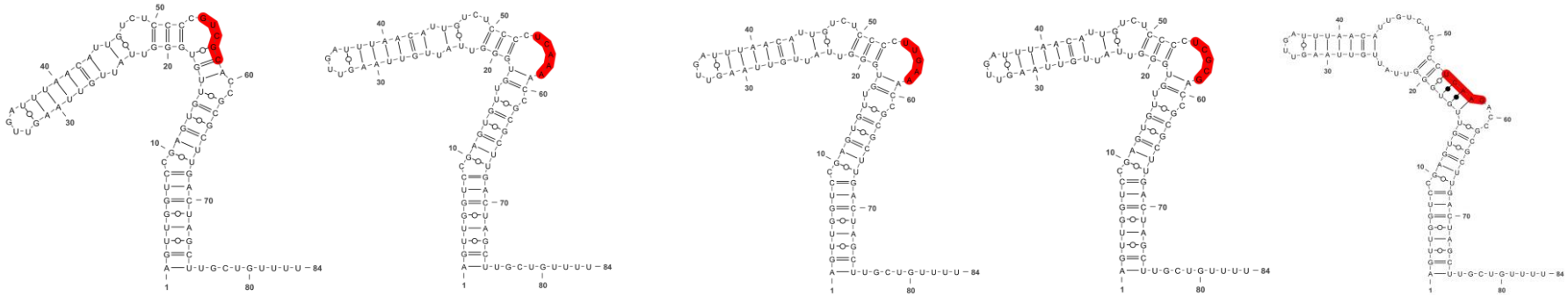


Figure 3.3. Some library 1 hY5 RNA mutants generate longer YsRNAs compared to mutant pool 2 and 3.

A. Size class distribution of YsRNAs from each hY5 library mutant pool replicate.

The graph shows the percentage of total reads for each sRNA cDNA library replicate of mutant pool 1, 2 and 3 at each size ranging from 25-34 nts. Mutant pool 2 and 3 YsRNAs (shown in green and blue) are mainly 32/33 nt. Mutant pool 1 YsRNA reads (shown in red) are longer with a length of 34/35 nt compared to mutant pool 2 and 3.

B. Abundance of library 1 hY5 RNA mutant sequences with the five most abundant YsRNAs.

The table shows the abundance of the L1 hY5 RNA mutant sequences with the five most abundant YsRNAs after bioinformatics analysis. All the L1 YsRNA library reads of each replicate were normalized and ranked by abundance. The first column shows the five nucleotide sequence motif of the L1 hY5 RNA mutant sequences that gave the most abundant YsRNA reads. In the second column the normalized abundance of each hY5 L1 sequence motif is shown.

C. Abundance of library 1 hY5 mutant sequences with the five least abundant YsRNAs

The table shows the abundance of the L1 hY5 RNA sequence motifs with the five least abundant YsRNAs after bioinformatics analysis. All the L1 YsRNA library reads of each replicate were normalized and ranked by abundance. The first column shows the five nucleotide sequence motif of the L1 hY5 RNA mutant sequences that resulted in the least abundant YsRNA reads. In the second column the normalized abundance of each hY5 L1 sequence motif is shown.

D. Predicted secondary structure of the wild type hY5 RNA.

The mutated region of L1 is shown in red. The hY5 RNA cleavage site is indicated with a red arrow.

E. Predicted secondary structure features of hY5 wild type RNA.

The features and structural elements of the predicted hY5 RNA secondary structure were adapted and modified after van Gelder *et al*., 1994 and Green *et al*., 1996. The examined secondary structure features were the stem S3, the internal loop L2a and the stem S2.

F. Predicted secondary structures of the library 1 hY5 RNA mutants with the five most abundant YsRNAs.

The mutation site of L1 hY5 mutants is shown in red. The hY5 L1 mutants that resulted in the five most abundant YsRNAs altered the secondary structure in a way that the cleavage site was 2 nt upshifted compared to the wild type hY5 RNA. Further, the cleavage occurred in both the wild type and mutant hY5 RNAs in the loop 2-3 nt away from the stem. The red arrows show the Y RNA cleavage site.

G. Predicted secondary structures of the library 1 hY5 RNA mutants with the five least abundant L1 hY5 RNA mutants.

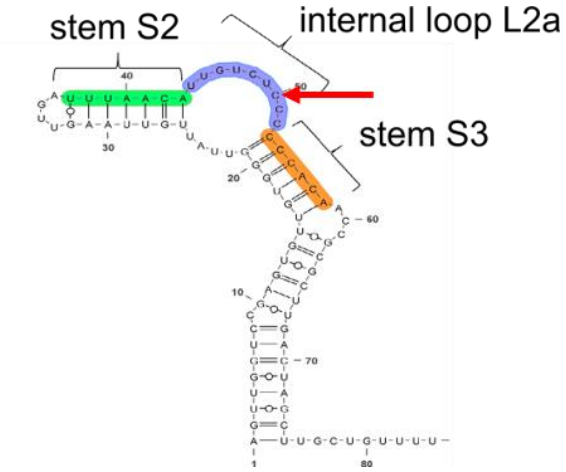
The mutation site of library 1 hY5 mutants is shown in red. The library 1 hY5 RNA mutants with the five least abundant YsRNAs generated 1 altered the secondary structure of the hY5 RNA in a way that much less or no Y RNAs were cleaved.

I analyzed the secondary structures of the library 1 hY5 RNA mutant sequences with the most abundant and least abundant YsRNAs in more detail (**Figure 3.4**).

When the secondary structures of the five most abundant and five least abundant library 1 hY5 mutants were analysed in more detail (see **Figure 3.4**, **B and C**) it was noted that in the library 1 hY5 mutants that generated the most abundant YsRNAs the stem S2 mainly consisted of 6 nt and had a high G-C content (**Figure 3.4**, **B**). However, the low abundant hY5 mutants folded in a way that the stem S2 was mostly only 3 nt or as seen in L1-10 longer but less G-C content and more A-U and G-U wobble base pairs (**Figure 3.4**, **C**).

This analysis suggested that the secondary structure features correlate with hY5 RNA cleavage. From the secondary structure analysis of the library 1 hY5 RNA mutants with the most abundant and least abundant YsRNAs it seemed that the internal loop and the stem S2 consisting of 5-6 nt and high G-C content might contribute to Y RNA cleavage from the 3' end. The internal loop of the most abundant hY5 library 1 mutants was mainly between 7-9 nt long and was pyrimidine rich, whereas the hY5 library 1 mutant that produced low amount of YsRNAs only had an internal loop of 3 nt. Notably, in the hY5 library 1 mutants that produced the most abundant hY5 YsRNAs the introduced mutations resulted in the change of the secondary structure in a way that the internal loop L2a got shortened by 1-2 nt compared to the wild type hY5 RNA. The stem S2 has less G-C base pairs and more A-U and G-U wobble base pairs which indicated that this region might be less stable and more flexible (**Figure 3.4**, **B**). These structural elements might be recognized by proteins/ribonuclease (s) that are involved in Y RNA cleavage from the 3' end of hY5 RNA.

A



B

RNA motif	Stem S3	Size of internal loop L2a	Stem S2
WT CCACA	7 nt 1x G-C 5x A-U 1x G-C	9 nt 4x U 2x G 3x C	6 nt 4x G-C 2x A-U
L1-1 ACGUG	7 nt 1x G-C 5x A-U 1x G-U	7 nt 4x U 1x G 2x C	6 nt 4x G-C 1x A-U 1x U-G
L1-2 ACCGC	7 nt 1x G-C 5x A-U 1x G-U	7 nt 4x U 1x G 2x C	12 nt 8x G-C 2x A-U 2x G-U
L1-3 CACAC	7nt 1x G-C 5x A-U 1x G-U	8 nt 4x U 1x G 3x C	6 nt 4x G-C 2x A-U
L1-4 AUGGC	7 nt 1x G-C 5x A-U 1x G-U	7 nt 4x U 2x G 1x C	12 nt 7x G-C 2x A-U 3x G-U
L1-5 ACGGC	7 nt 1x G-C 5x A-U 1x GU	7 nt 4x U 2x G 1x C	12 nt 8x G-C 2x A-U 1x G-U 1x U-bulge

C

RNA motif	Stem S3	Size of internal loop L2a	Stem S2
L1-6 GUCGC	10 nt 1x G-C 2x G-U 7x A-U	3 nt 2x U 1x C	3 nt 3x G-C
L1-7 UCAAA	10 nt 1x G-C 2x G-U 7x A-U	3 nt 2x U 1x C	3 nt 3x G-C
L1-8 UUGAA	10 nt 1x G-C 2x G-U 7x A-U	3 nt 2x U 1x C	3 nt 3x G-C
L1-9 UCGCG	10 nt 1x G-C 2x G-U 7x A-U	3 nt 2x U 1x C	3 nt 3x G-C
L1-10 UCAAG	7 nt 1x G-C 1x G-U 4x A-U	8 nt 4x U 1x G 1x C	6 nt 2x G-C 1x A-U 1x A-G

Figure 3.4. Secondary structure features of library 1 hY5 RNA mutants correlate with hY5 RNA cleavage.

A. Predicted secondary structure of wild type hY5 RNA.

The feature and structural elements of the predicted hY5 RNA secondary. The human Y5 RNA has a characteristic stem loop structure. The examined secondary structure features were the stem S3 (shown in orange), the internal loop L2a (shown in blue) and the stem S2 (shown in green). The 3' end hY5 RNA cleavage site is indicated with a red arrow.

B. Secondary structure features of library 1 hY5 RNA mutants with the most abundant YsRNAs.

The table gives an overview of the nucleotide and base pair composition of stem S3 (orange), internal loop L2a (blue) and stem S3 (green) for the library 1 hY5 RNA mutants with the five most abundant YsRNA reads.

C. Secondary structure features of library 1 hY5 RNA mutants with the least abundant YsRNAs.

The table gives an overview of the nucleotide and base pair composition of stem S3 (orange), internal loop L2a (blue) and stem S3 (green) for the library 1 hY5 RNA mutant sequences with the five least abundant YsRNA reads.

Interestingly, the sequencing data showed that the hY5 wild type motif (CCACA) was not the most abundant sequence motif among the longer 34-35 nt YsRNA reads, but was one of the most abundant sequence motifs among the shorter 32 nt YsRNA reads. This also indicated that the secondary structure seems to be involved in 3' end hY5 RNA cleavage.

Therefore, the results from the bioinformatics and secondary structure analysis further strengthened the hypothesis that the hY5 RNA cleavage at the 3' end correlates with the secondary structure, rather than the sequence.

After bioinformatics analysis the resulting YsRNA reads were normalized for all library replicates and sorted by their abundance. Therefore, I got a list of the most abundant and least abundant hY5 RNA sequence motifs that resulted in the most and least abundant mutant YsRNAs for hY5 RNA library 1. From this list, I selected the five most abundant and five least abundant mutant motifs of each library and cloned them as individual hY5 RNA mutants including the polymerase III promotor and terminator sequence into pGEMT easy vector. I transfected each individual mutant plasmid DNA and expressed the constructs in 3T3 mouse fibroblast cells. After 24 h I treated the cells with staurosporine in order to induce apoptosis and generate YsRNAs. After 8 h of treatment I extracted RNA and performed Northern blot analyses.

The aim of the Northern blot experiments was to compare and analyse the accumulation level of the individual YsRNAs derived from the mutant hY5 RNA sequences to the accumulation of YsRNAs derived from the wild type hY5 RNA (**Figure 3.5**). Intriguingly, the Northern blot analyses of the library 1 hY5 RNA that resulted in the five most abundant library 1 mutant YsRNAs (**Figure 3.5**, **A**) confirmed that the generated YsRNAs were longer compared to the wild type derived YsRNAs.

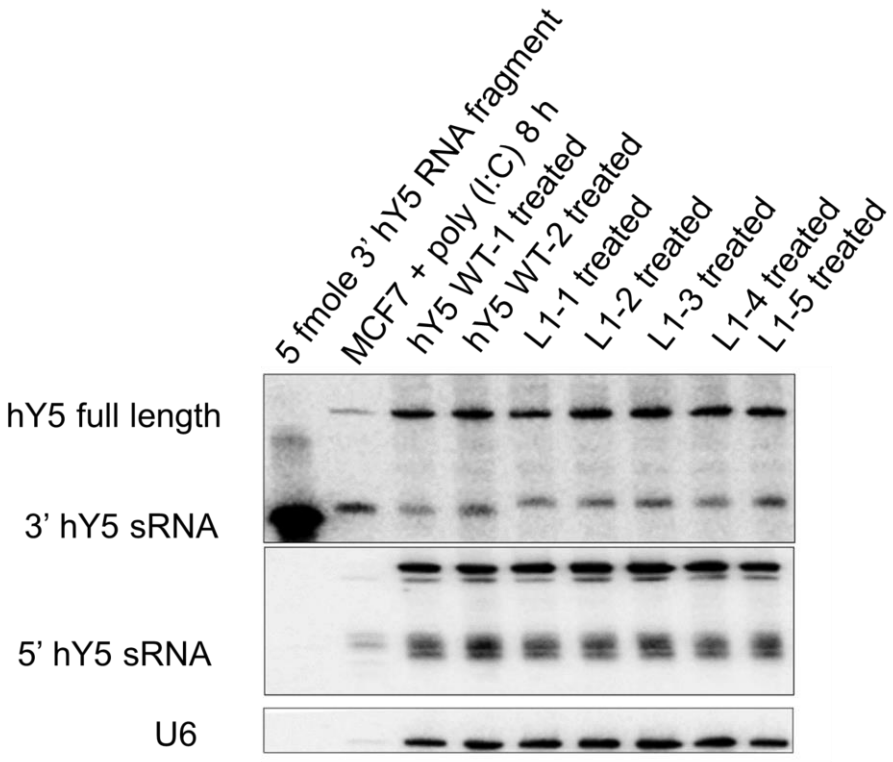
When the 3' end derived YsRNAs were quantified from the Northern blots by measuring the optical density and normalized to the loading control U6 and the wild type derived YsRNAs, I could observe that the wild type hY5 RNA generated the highest amount of YsRNAs compared to the other library 1 hY5 RNA mutant sequences that showed the most abundant YsRNAs in the bioinformatics analysis. However, the other L1 hY5 RNA mutants also produced a good amount of YsRNAs. From the relative quantification it can be concluded that all the library 1 hY5 mutant sequences with the most abundant YsRNAs were cleaved efficiently and resulted in a high amount of YsRNAs (**Figure 3.3**, **B**).

Interestingly, when I compared the ratios of YsRNAs were compared to the full length hY5 mutant RNA levels the library 1 hY5 RNA mutant L1-1 had a better ratio YsRNA:full length ratio than the wild type hY5 RNA. This would mean, that this library 1 hY5 RNA mutant might be processed more efficiently than the wild type hY5 RNA (**Figure 3.3**,**C**).

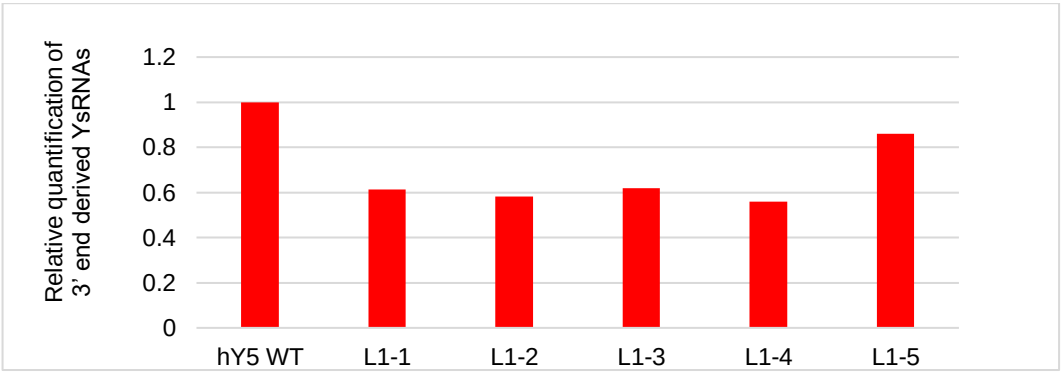
The library 1 hY5 RNA with the five least abundant YsRNA reads also showed in the relative quantification and ratio of full length hY5 mutant RNA sequence level to the YsRNA levels that these library 1 hY5 RNA mutants are cleaved at a much lower extent or not at all and cannot be efficiently processed to YsRNAs like the wild type hY5 RNA (**Figure 3.3**, **E and F**). These results further confirmed the sequencing results and bioinformatics analysis.

Secondary structure prediction (**Figure 3.3, F and G)** and Northern blot analysis of the library hY5 RNA mutants with the five most abundant and least abundant YsRNAs clearly showed that the hY5 RNA cleavage mainly occurs between the second and third nucleotide above the conserved stem. Therefore, these results confirm the results of the next generation sequencing and proved that the hY5 RNA cleavage from the 3' end correlates with the secondary structure.

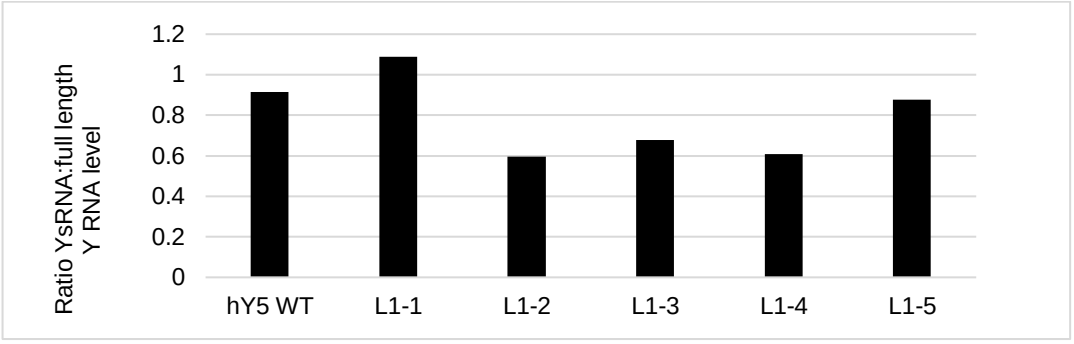
A


5 fmole 3' hY5 RNA fragment
MCF7 + poly (I:C) 8 h
hY5 WT-1 treated
hY5 WT-2 treated
L1-1 treated
L1-2 treated
L1-3 treated
L1-4 treated
L1-5 treated
hY5 full length
3' hY5 sRNA
5' hY5 sRNA
U6

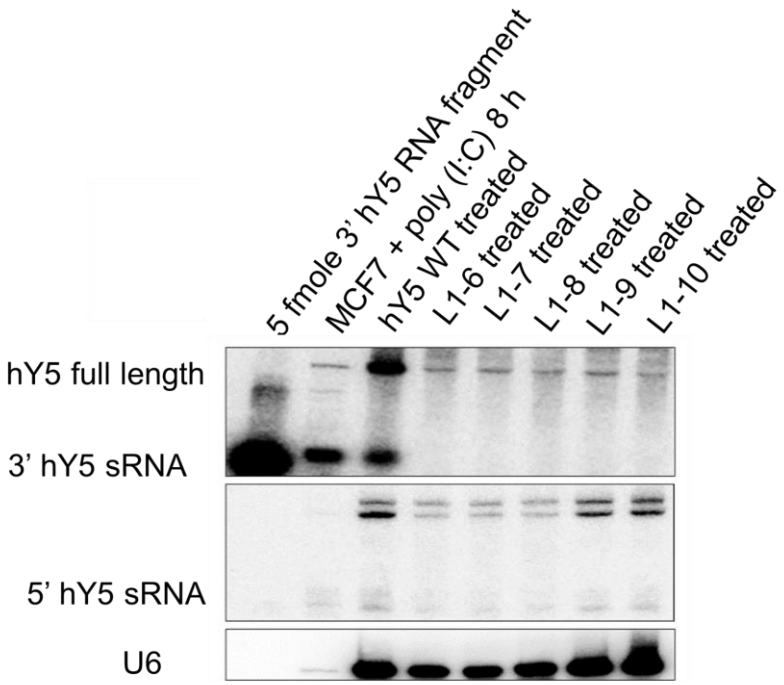
B


Relative quantification of 3' end derived YsRNAs
1.2
1
0.8
0.6
0.4
0.2
0
hY5 WT
L1-1
L1-2
L1-3
L1-4
L1-5

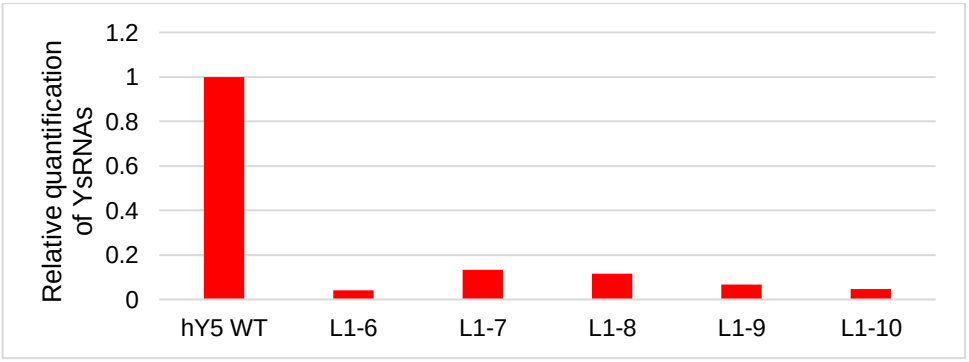
C


Ratio YsRNA:full length Y RNA level
1.2
1
0.8
0.6
0.4
0.2
0
hY5 WT
L1-1
L1-2
L1-3
L1-4
L1-5

D


5 fmole 3' hY5 RNA fragment
MCF7 + poly (I:C) 8 h
hY5 WT treated
L1-6 treated
L1-7 treated
L1-8 treated
L1-9 treated
L1-10 treated
hY5 full length
3' hY5 sRNA
5' hY5 sRNA
U6

E


Relative quantification of YsRNAs
1.2
1
0.8
0.6
0.4
0.2
0
hY5 WT
L1-6
L1-7
L1-8
L1-9
L1-10

F

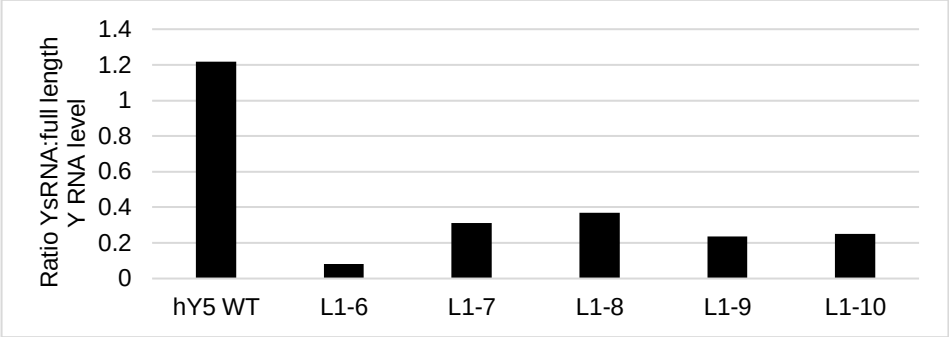


Figure 3.5. Most abundant library 1 hY5 RNA mutants generate longer YsRNAs.

A. Northern blot analysis of library 1 mutants with the five most abundant YsRNAs.

Total RNA extracted from human MCF7 cells treated with poly (I:C) and a synthetic 3' end derived hY5 RNA fragment with a length of 31 nt were used as size markers. The library 1 hY5 RNA mutants with the five most abundant YsRNAs were transfected into mouse cells and treated with staurosporine. The total RNA was extracted and Northern blot was performed. The Northern blot was probed with the 3' and 5' YsRNA probe. The blot was re-probed with U6 to demonstrate equal loading. The YsRNAs generated from the 3' end of the mutant hY5 RNAs were longer than the ones generated from wild type YsRNAs whereas the 5' end derived YsRNAs wild type had identical sizes for mutants and wild type.

B. Relative quantification of the most abundant 3' end derived mutant pool 1 YsRNAs.

The Northern blot of the most abundant mutant pool 1 YsRNAs was quantified using the ImageQuant software. The optical density of each mutant was normalized to the loading control U6 and the wild type hY5 RNA derived YsRNAs. The normalized optical density of the wild type derived YsRNAs was set at 1.

C. Calculated ratio between the expression level of full length hY5 mutant pool 1 hY5 RNAs compared to the expression level of most abundant mutant pool 1 YsRNAs.

The ratio of the normalized optical density of full length L1 hY5 RNA mutants and the derived YsRNAs was calculated and plotted on the graph.

D. Northern blot analysis of library 1 mutants with the five least abundant YsRNAs.

As a size marker for full length Y RNA and wild type sized Y RNA fragments, RNA extracted from human MCF7 cells treated with poly (I:C) and a synthetic 3' end derived hY5 RNA fragment with a length of 31 nt were used. The Northern blot was probed with the 3' end and 5' end YsRNA probe. The YsRNAs generated from the 3' end of the five least abundant library 1 mutants could not be detected by Northern blot. When the Northern blot was probed with the 5' end YsRNA probe a small amount of YsRNAs was observed. The blot was re-probed with U6 for equal loading.

E. Relative quantification of least abundant 3' end derived mutant pool 1 YsRNAs.

The Northern blot of the least abundant 3' end derived mutant pool 1 YsRNAs was quantified using the ImageQuant software. The optical density of each mutant was normalized to the loading control U6 and the wild type hY5 RNA derived YsRNAs. The normalized optical density of the wild type derived YsRNAs was set at 1.

F. Calculated ratio between the expression level of full length hY5 mutant pool 1 hY5 RNAs compared to the expression level of least abundant mutant pool 1 YsRNAs.

The ratio of the normalized optical density of full length L1 hY5 RNA mutants and the derived YsRNAs was calculated and plotted on the graph.

Interestingly, I could show by further secondary structure analysis of library 1 mutants that three other library 1 mutants had the same secondary structure than the wild type (**Figure 3.6**). When I looked at the sequencing reads derived from these library 1 mutants I could observe that they mainly produce YsRNAs of 32 nt like the wild type sequence (shown in red in **Figure 3.6, A**).

Among them, the hY5 library 1 mutants L1-CUACA and L1-UCACA showed wild type-sized YsRNAs after bioinformatics analysis and secondary structure folding (**Figure 3.6**, **B and C**). I cloned these mutants individually into pGEMT easy vector, transfected into mouse cells and treated with staurosporine to induce apoptosis and YsRNAs. After RNA extraction I performed Northern blot analysis and compared them to the two most abundant library 1 hY5 mutants L1-ACGUC and L1-ACCGC that generated longer YsRNAs. Remarkably, the Northern blot analysis clearly showed that the hY5 library 1 mutants L1-CUACA and L1-UCACA that folded the same way than wild type (**Figure 3.6**, **B**) generated wild type sized YsRNAs (**Figure 3.6**, **D**).

On the other hand the hY5 library 1 mutants L1-1 ACGUC and L1-2 ACCGC produced 1-2 nt longer YsRNAs (**Figure 3.3**, **F**, **Figure 3.6**, **C**) as previously shown as the mutations introduced in these hY5 RNA mutants changed the secondary structure in a way that the cleavage occurs more upstream (**Figure 3.6**, **D)**.

Interestingly, the library 1 hY5 RNA mutants L1-CUACA and L1-UCACA that had the same secondary structure than the wild type hY5 RNA showed a similar amount of YsRNAs in the relative quantification by optical densitometry. Intriguingly, the library 1 mutant L1-CUACA is processed to YsRNAs as efficient as the wild type hY5 RNA when the ratio of full length hY5 mutant RNA levels were compared to the YsRNA levels (**Figure 3.6**, **E and F**).

Surprisingly, in the L1 hY5 RNA mutant L1-ACCGC that generates longer YsRNAs than the wild type hY5 RNA the amount of YsRNAs generated is at the same level compared to the wild type hY5 RNA and is almost as efficiently processed to YsRNAs (**Figure 3.6**, **E and F**).

In conclusion, the random mutagenesis analysis suggests that the secondary structure might be more important than the nucleotide sequence to determine where the Y RNA cleavage occurs. Moreover, the Northern blot experiments confirmed that some of the mutations in library 1 alter the secondary structure in a way that the internal loop gets shorter and the Y RNA cleavage happens further upstream but still two nucleotides above the stem.

A

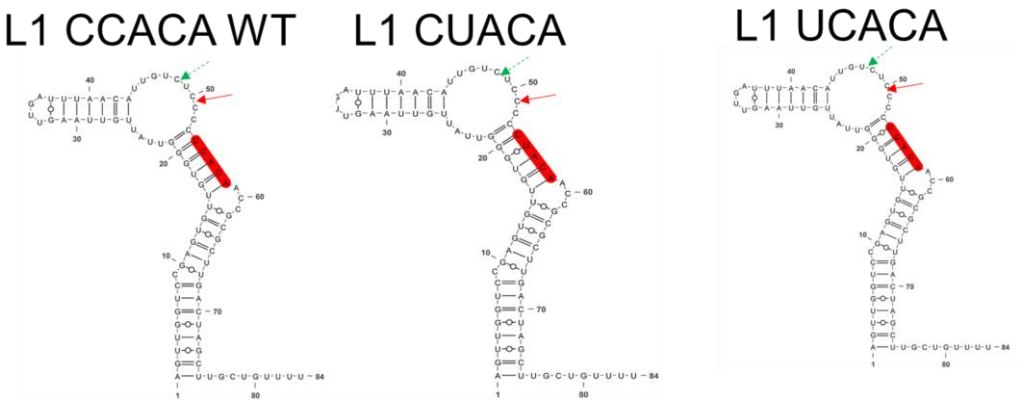
L1 32 nt YsRNA

RNA motif	Abundance
CCACA (WT)	892
CUACA	365
UCACA	215
ACGUC	35
ACCGC	24

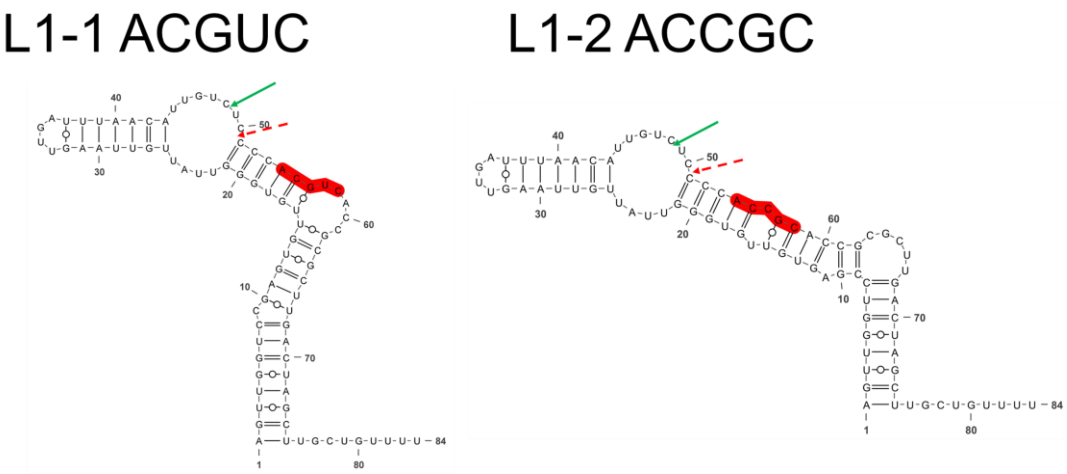
L1 34 nt YsRNA

RNA motif	Abundance
CCACA	424
CUACA	8
UCACA	8
ACGUC	3430
ACCGC	2722

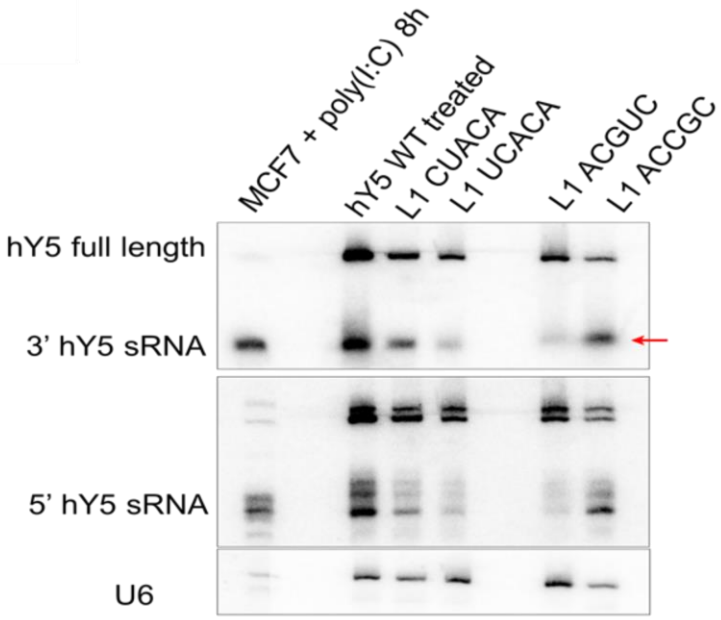
B



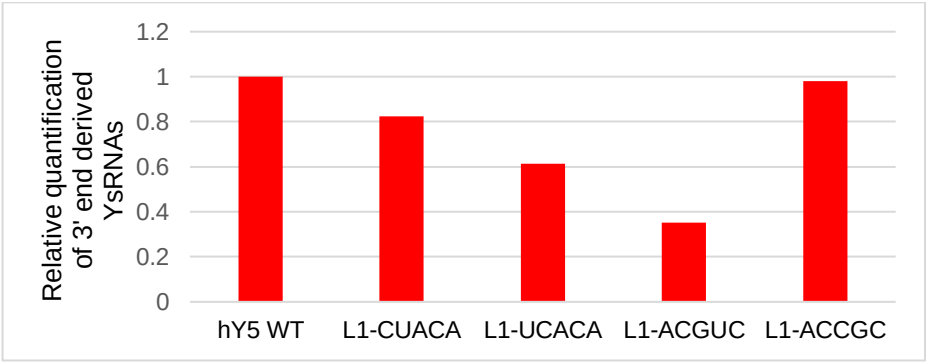
C



D


MCF7 + poly(I:C) 8h
hY5 WT treated
L1 CUACA
L1 UCACA
L1 ACGUC
L1 ACCGC
hY5 full length
3' hY5 sRNA
5' hY5 sRNA
U6

E


Relative quantification of 3' end derived YsRNAs
1.2
1
0.8
0.6
0.4
0.2
0
hY5 WT
L1-CUACA
L1-UCACA
L1-ACGUC
L1-ACCGC

F

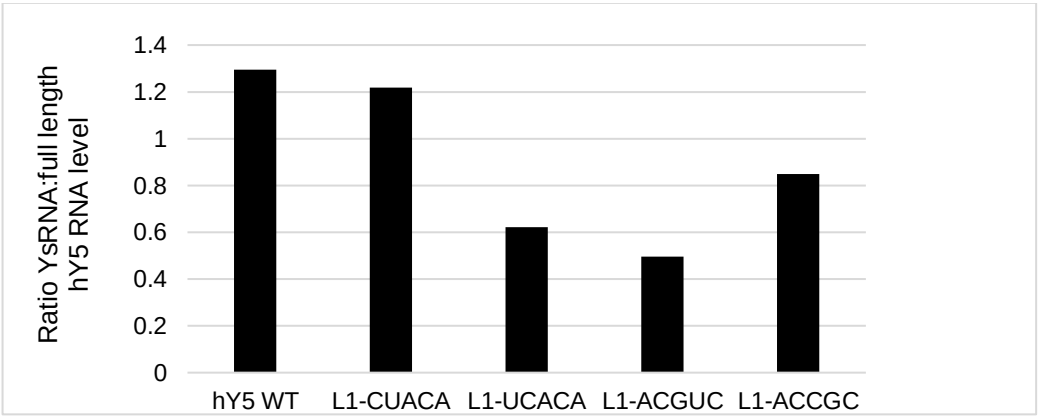

Ratio YsRNA:full length hY5 RNA level
1.4
1.2
1
0.8
0.6
0.4
0.2
0
hY5 WT
L1-CUACA
L1-UCACA
L1-ACGUC
L1-ACCGC

Figure 3.6. Library 1 hY5 mutants with exactly the same secondary structure as wild type produce wild type sized YsRNAs compared to the most abundant hY5 library 1 mutants that produce longer YsRNAs.

A. Abundances of wild type hY5 derived YsRNA reads and two hY5 library 1 mutants with the same secondary structure as wild type hY5 RNA and 32 nt YsRNA reads and two most abundant hY5 library 1 mutants that generate the most 34 nt YsRNA reads.

The table shows the normalized abundances of wild type hY5 RNA and the YsRNA reads of the mutants L1-CUACA, L1-UCACA, L1-ACGUC and L1-ACCGC normalized across three replicates with sizes of 32 and 34 nts.

B. Predicted secondary structures of wild type hY5 RNA and two hY5 library 1 mutants with the same secondary structure as the wild type hY5 RNA.

The mutants L1-CUACA and L1-UCACA have exactly the same secondary structure than the wild type hY5 RNA. The mutated region is indicated in red. The red arrows show the Y RNA cleavage site resulting in YsRNAs with 32 nts. The arrow in green indicates the Y RNA cleavage site that leads to YsRNAs with a size of 34 nts.

C. Predicted secondary structures of wild type hY5 RNA and two hY5 library 1 mutants that generated longer YsRNAs.

In the hY5 library 1 mutants L1-ACGUC and L1-ACCGC the secondary structure is changed in a way that the Y RNA cleavage occurs more upstream. The mutated region is indicated in red. The red arrows show the Y RNA cleavage site resulting in YsRNAs with 32 nts. The arrow in green indicates the Y RNA cleavage site that leads to YsRNAs with a size of 34 nts.

D. Library 1 hY5 RNA mutants with exactly the same secondary structure as wild type produce wild type sized YsRNAs compared to the most abundant hY5 library 1 mutants that produce longer YsRNAs.

For Northern blot analysis total RNA extracted from human MCF7 cells treated with poly (I:C) was used as a size marker. Two library 1 mutants that folded the same way than the wild type hY5 RNA and the two most abundant library 1 mutants that generated longer YsRNAs were transfected into mouse cells and treated with staurosporine. The total RNA was extracted and Northern blot was performed. The Northern blot was probed with the 3' and 5' YsRNA probe. The blot was re-probed with U6 to demonstrate equal loading. The YsRNAs generated from the 3' end of the two most abundant mutant hY5 RNAs (L1-ACGUC, L1-ACCGC) were longer than the ones generated from wild type YsRNAs and the two tested hY5 library 1 mutants (L1-CUACA, L1-UCACA) with the same secondary structure than the wild type. The 5' end derived YsRNAs had identical sizes for mutants and wild type.

E. Relative quantification of 3' end derived mutant pool 1 YsRNAs.

The Northern blot of 3' end derived mutant pool 1 YsRNAs was quantified using the ImageQuant software. The optical density of each mutant was normalized to the loading control U6 and the wild type hY5 RNA derived YsRNAs. The normalized optical density of the wild type derived YsRNAs was set at 1.

F. Calculated ratio between the expression level of full length hY5 mutant pool 1 hY5 RNAs compared to the expression level of most abundant mutant pool 1 YsRNAs.

The ratio of the normalized optical density of full length L1 hY5 RNA mutants and the derived YsRNAs was calculated and plotted on the graph.

3.2.3 Library 2 hY5 RNA mutants generate YsRNAs at the same size as wild type

After bioinformatics analysis the YsRNAs derived from library 2 were ranked by abundance (**Figure 3.7**, **A**). The library 2 hY5 RNA mutants with the top five most abundant and five least abundant YsRNAs from each mutant pool were chosen for experimental validation and cloned individually into the expression vector. I confirmed the sequences of the constructs by Sanger sequencing. I transfected the individual hY5 mutants into 3T3 cells and I tested the accumulation level of YsRNAs after staurosporine treatment by Northern blot (**Figure 3.7**, **E**).

I predicted the secondary structures of 30 library 2 hY5 RNA mutants with the most abundant and least abundant YsRNAs using RNAfold. For library 2 mutants it can be seen that that all the secondary structures look similar (**Figure 3.7**, **C and D**) to the wild type hY5 RNA (**Figure 3.7, B**) and that the cleavage occurred at the wild type hY5 RNA cleavage site (**Figure 3.7**, **E**).

The library 2 hY5 RNA mutants that resulted in the top five most abundant and five least abundant YsRNAs (**Figure 3.7**, **A)** were transfected into mouse cells and experimental validation was performed by testing the expression level and hY5 RNA cleavage pattern by Northern blot (**Figure 3.7**, **E and G***)*.

The Northern blot analysis of the library 2 hY5 RNA mutants that generate the five most abundant YsRNAs (**Figure 3.7**, **E**) confirmed that the hY5 RNA cleavage mainly occurred between positions C50 and C51 as seen for the wild type hY5 RNA. The Northern blots and relative quantification also showed that the YsRNAs derived from the 3' end of the five least abundant mutants were generated at a lower amount as seen in the next generation sequencing (**Figure 3.7**, **G and H**). The relative quantification of the most abundant library 2 hY5 RNA mutant derived YsRNAs to wild type derived YsRNAs by densitometry showed that the YsRNA levels of the most abundant library 2 hY5 RNA mutant sequences were similar to the amount of wild type hY5 derived YsRNAs (**Figure 3.7**, **F**). Thus, the region that was mutated in library 2 hY5 RNA mutants might be more flexible and not be crucial for efficient Y RNA processing.

A

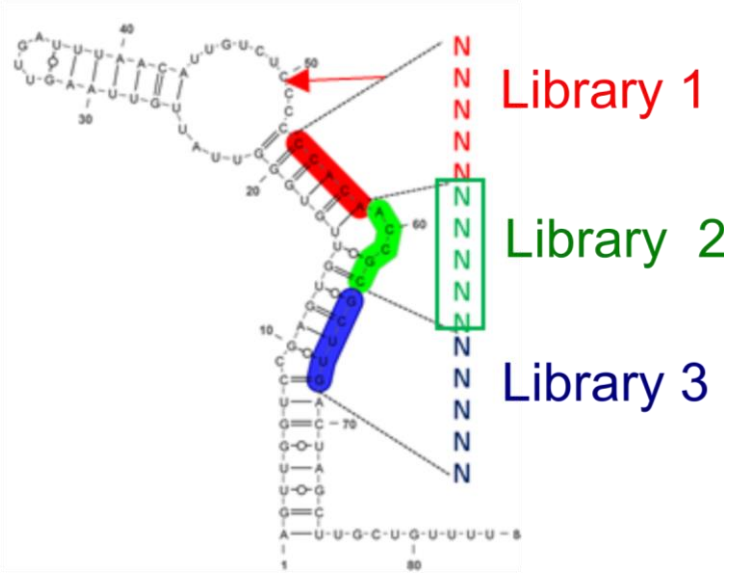
Five most abundant motifs

Mutant	Abundance
L2-1 CCGUU	4027
L2-2 AGGUG	2865
L2-3 CCGUC	2236
L2-4 CUGUC	1908
L2-5 CUGUG	1902

Five least abundant motifs

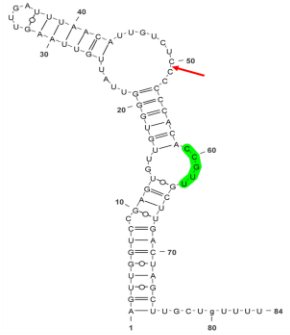
Mutant	Abundance
L2-6 GUCCG	156
L2-7 UCAAA	11
L2-8 UUGAA	2
L2-9 UCGCG	1
L2-10 UCAAG	1

B

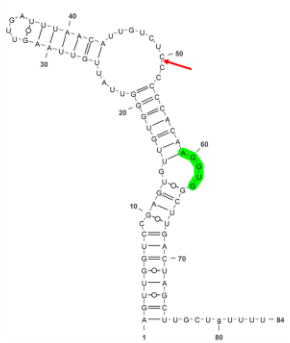


C

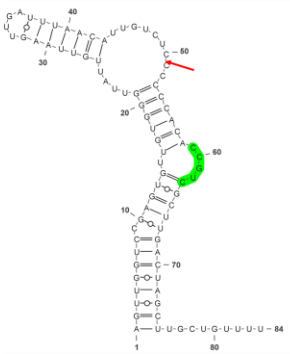
L2-1 CCGUU



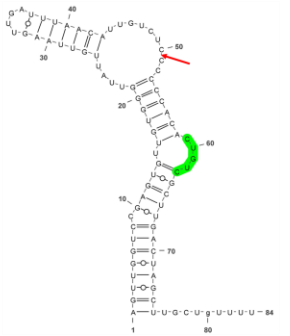
L2-2 AGGUG



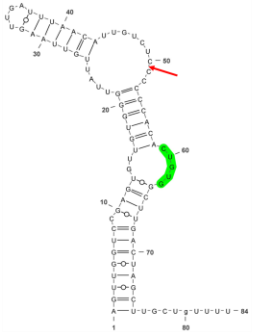
L2-3 CCGUC



L2-4 CUGUC

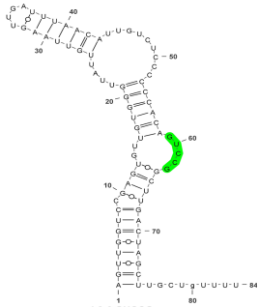


L2-5 CUGUG

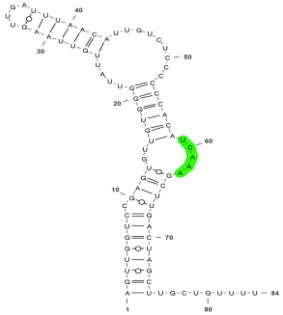


D

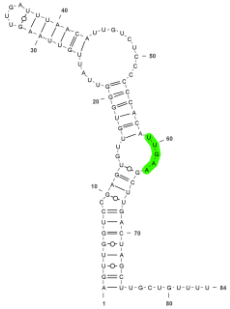
L2-6 GUCCG



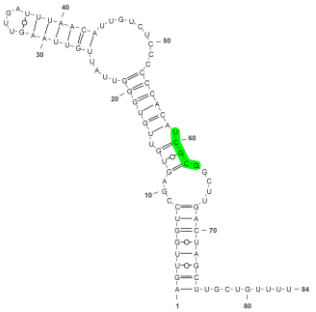
L2-7 UCAAA



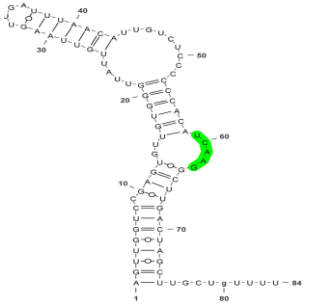
L2-8 UUGAA



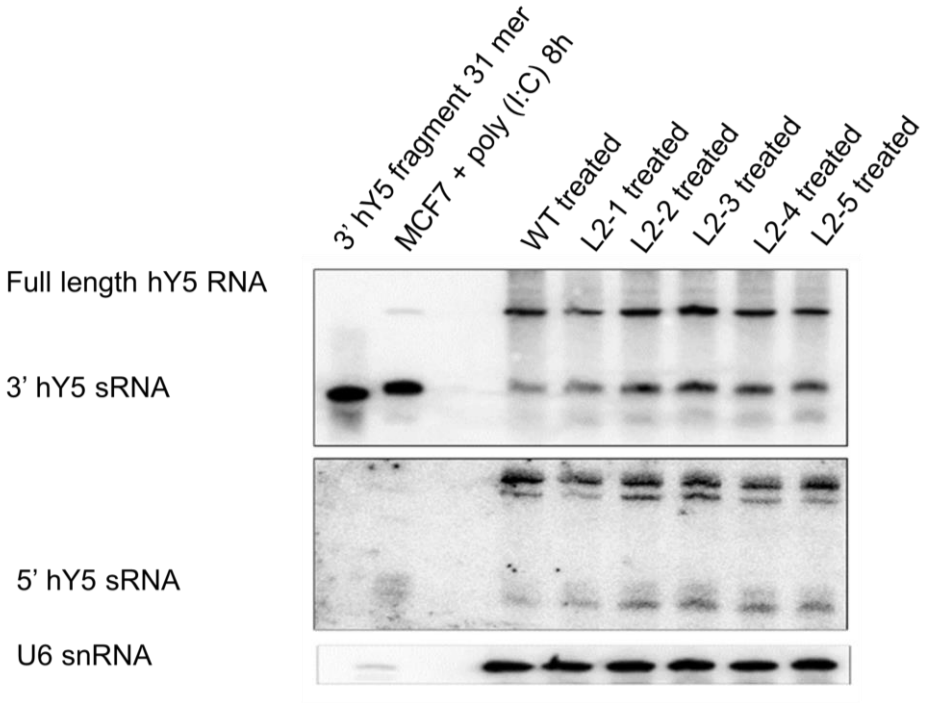
L2-9 UCGCG



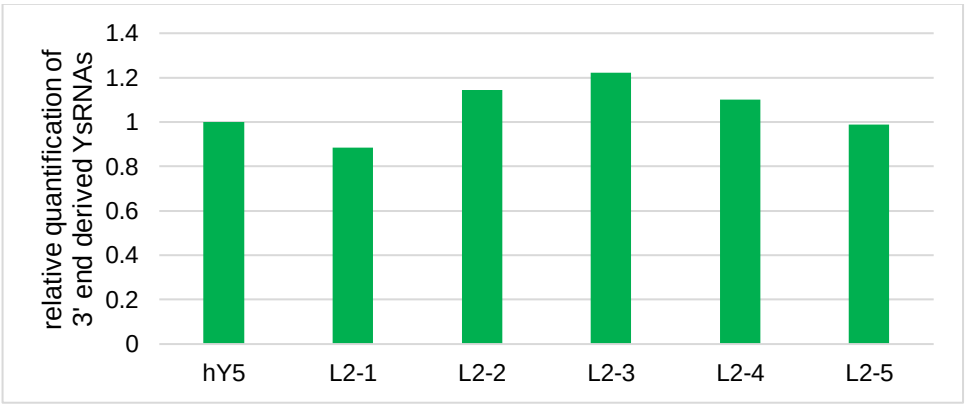
L2-10 UCAAG



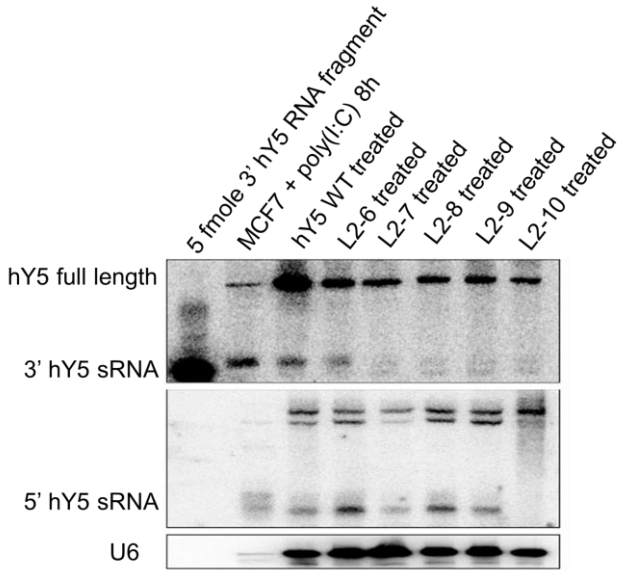
E


3' hY5 fragment 31 mer
MCF7 + poly (I:C) 8h
WT treated
L2-1 treated
L2-2 treated
L2-3 treated
L2-4 treated
L2-5 treated
Full length hY5 RNA
3' hY5 sRNA
5' hY5 sRNA
U6 snRNA

F


relative quantification of 3' end derived YsRNAs
1.4
1.2
1
0.8
0.6
0.4
0.2
0
hY5
L2-1
L2-2
L2-3
L2-4
L2-5

G


5 fmole 3' hY5 RNA fragment
MCF7 + poly(I:C) 8h
hY5 WT treated
L2-6 treated
L2-7 treated
L2-8 treated
L2-9 treated
L2-10 treated
hY5 full length
3' hY5 sRNA
5' hY5 sRNA
U6

H

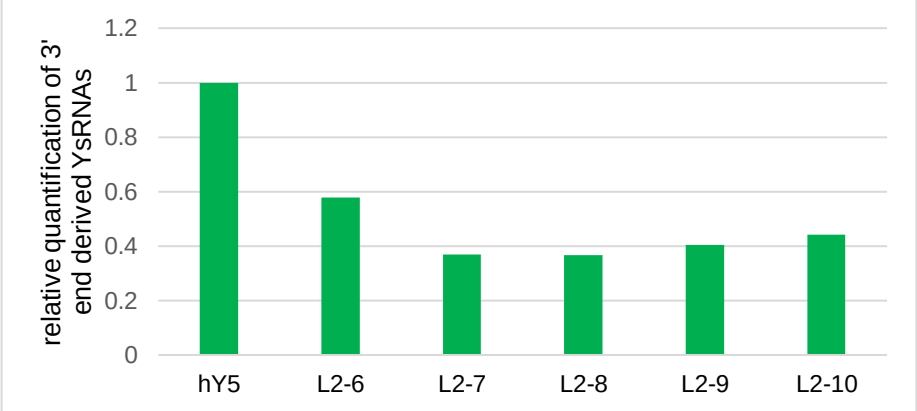

relative quantification of 3' end derived YsRNAs
1.2
1
0.8
0.6
0.4
0.2
0
hY5
L2-6
L2-7
L2-8
L2-9
L2-10

Figure 3.7. Library 2 hY5 RNA mutants produce YsRNAs at the same size as wild type.

A. Abundances of library 2 hY5 RNA mutants with the five most abundant and five least abundant YsRNA reads.

The table shows the abundance of L2 hY5 RNA mutant sequences with the five most abundant and five least abundant YsRNAs after bioinformatics analysis. All L2 YsRNA library reads of each replicate were normalized and ranked by abundance. The first column shows the five nucleotide sequence motif of L2 hY5 RNA mutant sequences that gave the most abundant/least abundant YsRNA reads. In the second column the normalized abundance of each L2 hY5 RNA sequence motif is shown.

B. Predicted secondary structure of the wild type hY5 RNA.

The secondary structure was predicted using RNAfold and visualized using Varna. The region mutated in L2 mutants is highlighted in green.

C. Predicted secondary structures of library 2 hY5 RNA mutants with the five most abundant YsRNAs.

The secondary structures were predicted using RNAfold. The library 2 hY5 RNA mutants with the five most abundant YsRNAs fold similar to the wild type hY5 RNA. The red arrows show the Y RNA cleavage site.

D. Predicted secondary structures of library 2 hY5 RNA mutants with the five least abundant YsRNAs.

Predicted secondary structures of library 2 hY5 RNA mutants with the five least abundant YsRNAs RNAfold. The secondary structures of library 2 hY5 RNA mutants with the five least abundant YsRNAs are similar to the wild type hY5 RNA structure.

E. Northern blot analysis of library 2 hY5 RNA mutants with the five most abundant YsRNAs.

Total RNA extracted from human MCF7 cells treated with poly (I:C) and a synthetic 3' end derived hY5 fragment of 31 nt were used as size markers for full length hY5 RNA and wild type-sized YsRNAs. The total RNAs extracted from the five most abundant library 2 hY5 mutants that were transfected into mouse cells and treated with staurosporine were analysed by Northern blotting. The Northern blot was probed with the 3' and 5' YsRNA probe. The blot was re-probed with U6 as a loading control. The YsRNAs produced from the five most abundant hY5 library 2 mutants were at the same size as wild type hY5 YsRNAs both from the 3' end and the 5' end of the hY5 RNA.

F. Relative quantification of the most abundant 3' end derived mutant pool 2 YsRNAs.

The Northern blot of the most abundant mutant pool 2 YsRNAs was quantified using the ImageQuant software. The optical density of each mutant was normalized to the loading control U6 and the wild type hY5 RNA derived YsRNAs. The normalized optical density of the wild type derived YsRNAs was set at 1.

G. Northern blot analysis of library 2 hY5 RNA mutants with the five least abundant YsRNAs.

As a size marker for full length Y RNA and wild type sized Y RNA fragments RNA from human MCF7 cells treated with poly (I:C) and a synthetic 3' end derived hY5 RNA fragment with a length of 31 nt were used. After transfection of the least abundant hY5 library 2 mutant plasmids into mouse cells and treatment with staurosporine the RNA was subjected to Northern blot analysis. The Northern blot was probed with the 3' and 5' YsRNA probe. For equal loading the Northern blot was re-probed with U6. The five least abundant hY5 library 2 mutants generated wild type-sized YsRNA but at a lower level than wild type hY5 RNA.

H. Relative quantification of the least abundant 3' end derived mutant pool 2 YsRNAs.

The Northern blot of the least abundant mutant pool 2 YsRNAs was quantified using the ImageQuant software. The optical density of each mutant was normalized to the loading control U6 and the wild type hY5 RNA derived YsRNAs. The normalized optical density of the wild type derived YsRNAs was set at 1.

3.2.4 Library 3 hY5 RNA mutants generate YsRNAs at the same size as wild type hY5 RNA

After bioinformatics analysis and secondary structure prediction of library 3 hY5 RNA mutants with the 30 most abundant and least abundant library 3 hY5 RNA YsRNAs I could observe that the secondary structures of library 3 mutants folded in a similar way as the wild type hY5 RNA (**Figure 3.8, C and D**).

The library 3 hY5 RNA mutants with the five most abundant and five least abundant YsRNA reads (**Figure 3.8**, **A**) were cloned individually as previously described, transfected into mouse cells and experimental validation was performed by testing the expression level and hY5 RNA cleavage pattern by Northern blot (**Figure 3.8, E and G**).

The Northern blot analysis and relative quantification of the five most abundant (**Figure 3.8, E and F**) and five least abundant hY5 library 3 mutants (**Figure 3.8, G and H**) demonstrated that the hY5 RNA cleavage mainly occurred at the position C50 and C51 and therefore mutant pool 3 hY5 RNAs generate wild type-sized YsRNAs.

The relative quantification of the Northern blots with the most abundant and least abundant 3' end derived mutant pool 3 YsRNAs compared to the wild type hY5 RNA derived YsRNAs showed that the levels of YsRNAs generated by the library 3 hY5 RNA mutants were similar to the wild type hY5 RNA (**Figure 3.8**, **F and H**). This indicated that the mutations introduced in this region might not be that important for YsRNA generation and is more flexible compared to the mutated region of hY5 library 1. However, when the Ro60 binding site is mutated no Y RNA can occur because the hY5 mutant sequences are not stable anymore and undergo degradation (**Figure 3.2**).

A

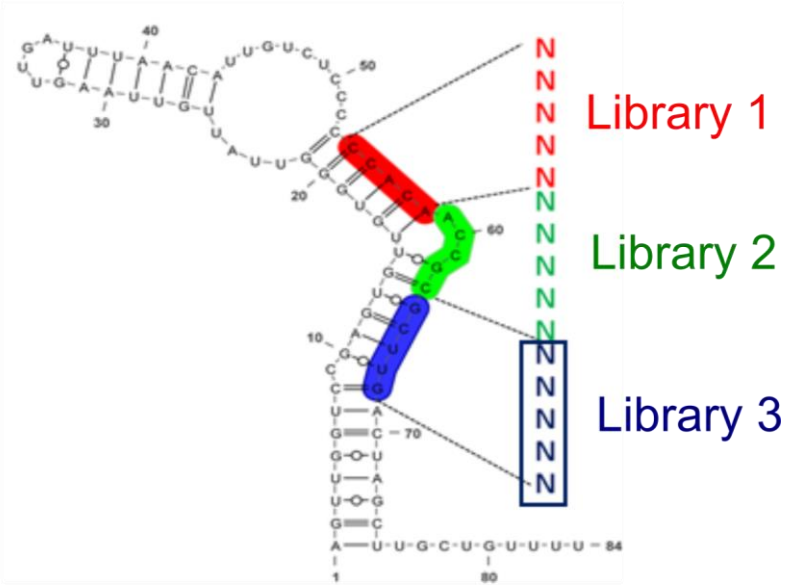
Five most abundant motifs

Mutant	Abundance
L3-1 GUUUG	16755
L3-2 UGUAG	9742
L3-3 GUUCG	4523
L3-4 CGUUG	3066
L3-5 CGUCG	1441

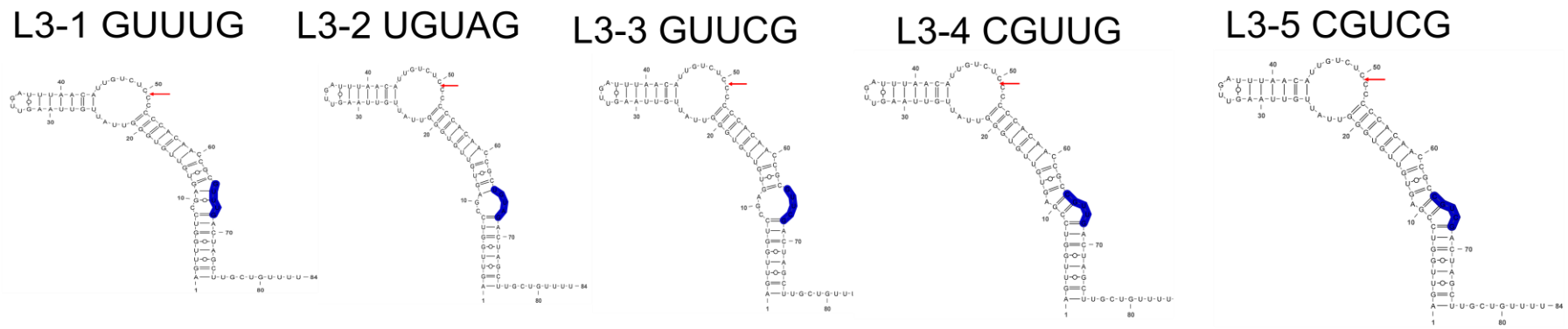
Five least abundant motifs

Mutant	Abundance
L3-6 UCCCA	1
L3-7 UAAUU	1
L3-8 GAAAC	1
L3-9 AUCUA	1
L3-10 CGACC	1

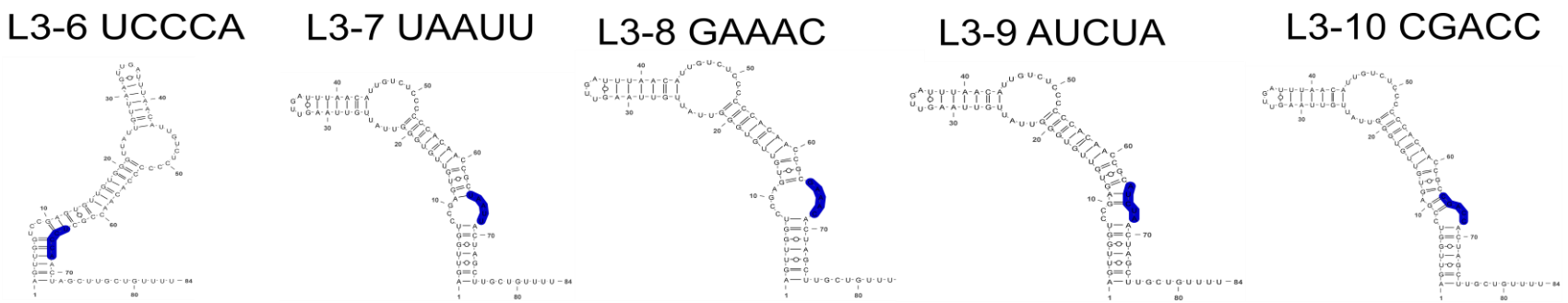
B



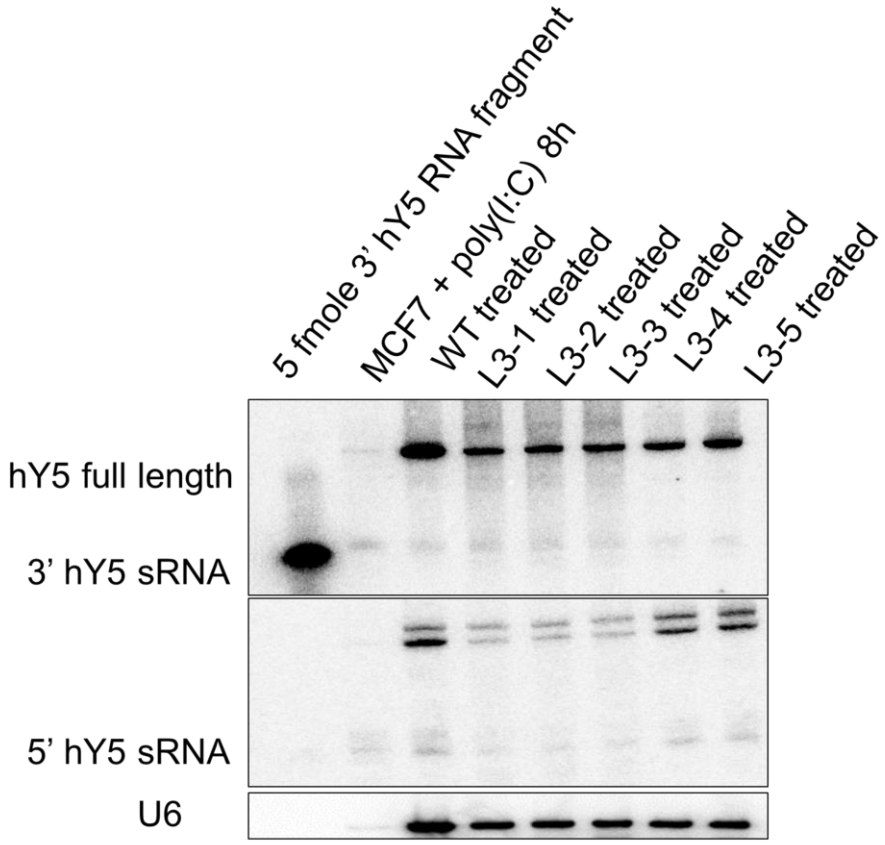
C


L3-1 GUUUG
L3-2 UGUAG
L3-3 GUUCG
L3-4 CGUUG
L3-5 CGUCG

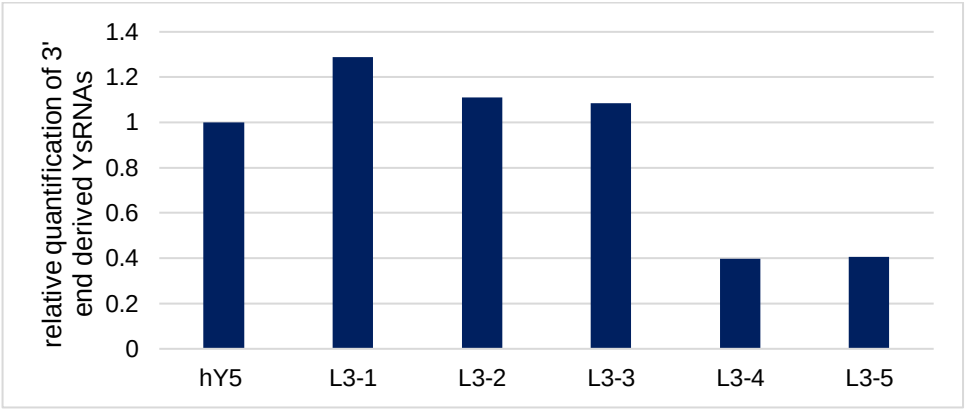
D


L3-6 UCCCA
L3-7 UAAUU
L3-8 GAAAC
L3-9 AUCUA
L3-10 CGACC

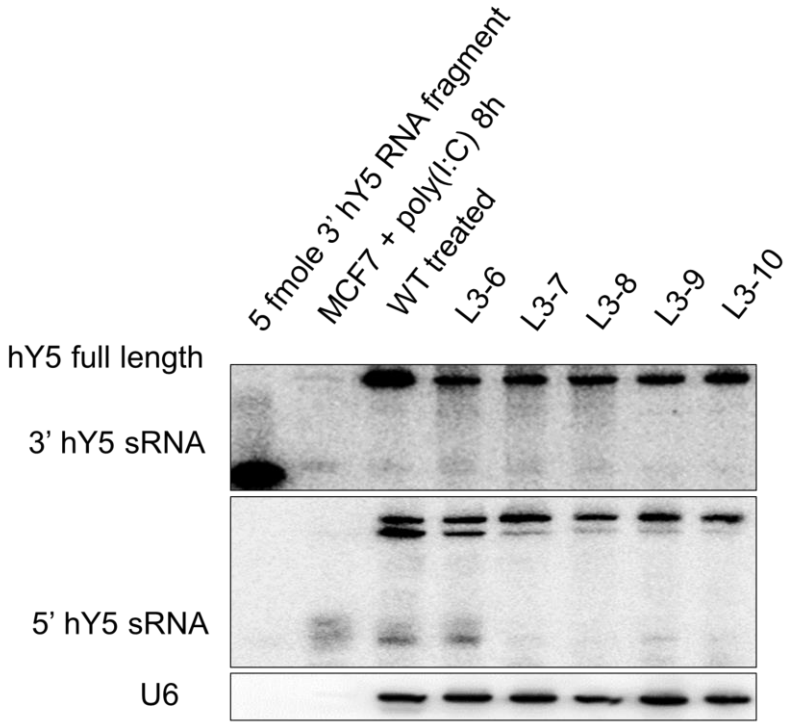
E


5 fmole 3' hY5 RNA fragment
MCF7 + poly(I:C) 8h
WT treated
L3-1 treated
L3-2 treated
L3-3 treated
L3-4 treated
L3-5 treated
hY5 full length
3' hY5 sRNA
5' hY5 sRNA
U6

F


relative quantification of 3' end derived YsRNAs
1.4
1.2
1
0.8
0.6
0.4
0.2
0
hY5
L3-1
L3-2
L3-3
L3-4
L3-5

G


5 fmole 3' hY5 RNA fragment
MCF7 + poly(I:C) 8h
WT treated
L3-6
L3-7
L3-8
L3-9
L3-10
hY5 full length
3' hY5 sRNA
5' hY5 sRNA
U6

H

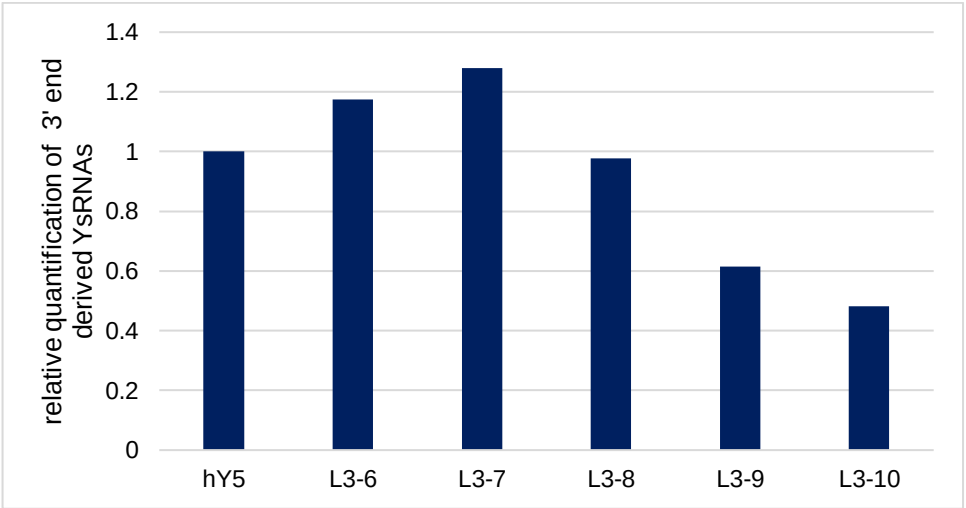

relative quantification of 3' end derived YsRNAs
1.4
1.2
1
0.8
0.6
0.4
0.2
0
hY5
L3-6
L3-7
L3-8
L3-9
L3-10

Figure 3.8. Library 3 hY5 RNA mutants generate YsRNAs at the same size as wild type hY5 RNA.

A. Abundances of library 3 hY5 RNA mutants with the five most abundant and five least abundant YsRNA reads.

The table shows the abundance of library 3 hY5 RNA mutant sequences with the five most abundant and five least abundant YsRNAs after bioinformatics analysis. All L3 YsRNA library reads of each replicate were normalized and ranked by abundance. The first column shows the five nucleotide sequence motif of L3 hY5 RNA mutant sequences that gave the most abundant/least abundant YsRNA reads. In the second column the normalized abundance of each L2 hY5 RNA sequence motif is shown.

B. Predicted secondary structure of the wild type hY5 RNA.

The secondary structure was predicted using RNAfold and visualized using Varna. The region mutated in L3 hY5 RNA mutants is highlighted in dark blue.

C. Predicted secondary structures of library 3 hY5 RNA mutants with the five most abundant YsRNAs.

The secondary structures were predicted using RNAfold. The secondary structures of L3 hY5 RNA mutants with the most abundant YsRNAs fold similar to the wild type hY5 RNA. The red arrows show the Y RNA cleavage site. The Y RNA cleavage occurred between C50 and C51.The red arrows show the Y RNA cleavage site.

D. Predicted secondary structures of library 2 hY5 RNA mutants with the five least abundant YsRNAs.

Predicted secondary structures of L3 hY5 RNA mutants with the five least abundant YsRNAs using RNAfold. The secondary structures of the L3 hY5 RNA mutant sequences that have the least abundant YsRNA reads fold in a similar way than the wild type hY5 RNA sequence.

E. Northern blot analysis of library 3 hY5 RNA mutants with the five most abundant YsRNAs.

Total RNA extracted from human MCF7 cells treated with poly (I:C) and a synthetic 3' end derived hY5 fragment of 31 nt were used as size markers for full length hY5 RNA and wild type-sized YsRNAs. The five most abundant library 3 mutants were transfected into mouse cells and YsRNA production was induced by staurosporine. The total RNA was extracted and analysed by Northern blotting. The Northern blot was probed with the 3' end and 5' end hY5 RNA probe. The blot was re-probed with U6 to demonstrate equal loading. The five most abundant hY5 library 3 mutants produced wild type-sized YsRNAs from the 3' and 5' end of the hY5 RNA.

F. Relative quantification of the most abundant 3' end derived mutant pool 3 YsRNAs.

The Northern blot of the most abundant mutant pool 3 YsRNAs was quantified using the ImageQuant software. The optical density of each mutant was normalized to the loading control U6 and the wild type hY5 RNA derived YsRNAs. The normalized optical density of the wild type derived YsRNAs was set at 1.

G. Northern blot analysis of library 3 hY5 RNA mutants with the five least abundant YsRNAs.

As a size marker for full length Y RNA and wild type sized Y RNA fragments RNA from human MCF7 cells treated with poly (I:C) and a synthetic 3' end derived hY5 RNA fragment with a length of 31 nt were used. The hY5 library 3 mutant sequences that generated the five least abundant YsRNA reads were transfected into mouse cells and the expression level and YsRNA production was measured by Northern blot. The Northern blot was probed with the 3' and 5' end hY5 RNA probe. For equal loading the Northern blot was re-probed with U6. The hY5 RNA library 3 mutants with the five least abundant YsRNA reads produced less YsRNAs compared to the wild type.

H. Relative quantification of the least abundant 3' end derived mutant pool 3 YsRNAs.

The Northern blot of the least abundant mutant pool 3 YsRNAs was quantified using the ImageQuant software. The optical density of each mutant was normalized to the loading control U6 and the wild type hY5 RNA derived YsRNAs. The normalized optical density of the wild type derived YsRNAs was set at 1.

3.2.5 A bulge of one nucleotide is required for human Y5 RNA cleavage

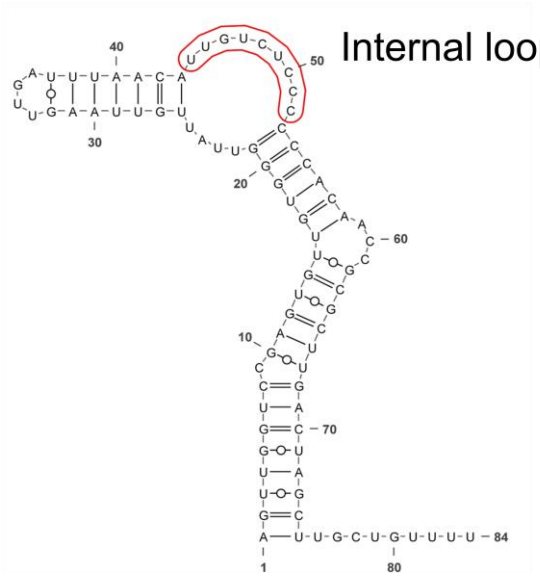
The aim of this part of the work was to explore if the length of the loop L2a (**Figure 3.9**, **A,** shown in red) is important for human Y5 RNA cleavage. Therefore, I designed and generated nine hY5 RNA mutants in order to sequentially delete the loop from the 5' end of the hY5 RNA. These hY5 RNA mutants were termed L2a loop mutants.

Secondary structure analysis of L2a loop mutants showed that all the deletion mutants except L2a Δ 8nt and L2a Δ 9nt fold similar to the wild type hY5 RNA (**Figure 3.9**, **B**). I transfected all the L2a loop deletion mutants into mouse cells and treated the cells with staurosporine to induce YsRNA generation. The RNAs were subjected to Northern blot analysis and the accumulation of YsRNA in cells transfected with the deletion mutants was compared to the wild type hY5 RNA (**Figure 3.9**, **C, D and E**).

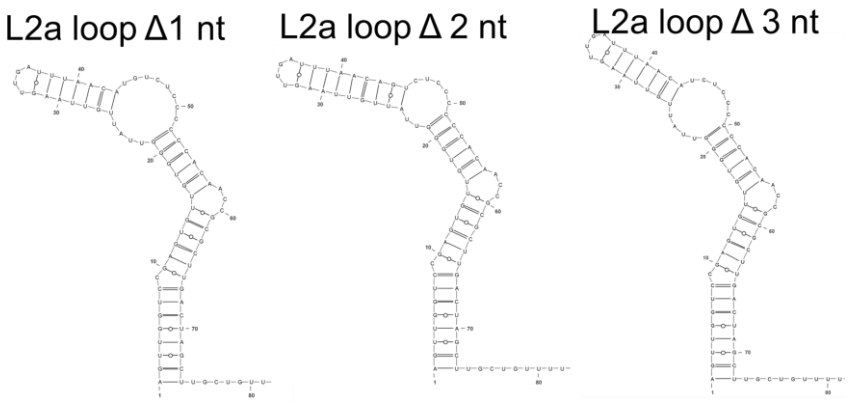
Northern blot analysis and relative quantification showed that sequential deletion of the loop L2a only had a small effect on YsRNA production from either the 3' or the 5' end of the hY5 RNA compared to the wild type hY5 derived YsRNAs (**Figure 3.9**, **C and D**). The mutants L2a loop Δ 1 nt to L2a loop Δ 8 nt showed decreased levels of YsRNAs compared to the wild type hY5 RNA. The mutant L2a loop Δ 9nt did not produce YsRNAs either from the 3' nor the 5' end of hY5. Also, it has to be noted that the full length L2a loop Δ 9 nt RNA could not be detected because the mutant RNA might not be stable (**Figure 3.9**, **C and D**). Interestingly, the relative quantification of L2a loop derived YsRNAs showed that the mutant L2a loop Δ 4 nt, L2a loop Δ 5 nt and L2a loop Δ 6 nt generated more YsRNAs as the wild type hY5 RNA (**Figure 3.9**, C). This could be explained due to the fact that these mutants fold similar to the wild type hY5 RNA (**Figure 3.9**, **B**) and it could be that they are more accessible to RNases to process these hY5 RNA L2a loop mutants.

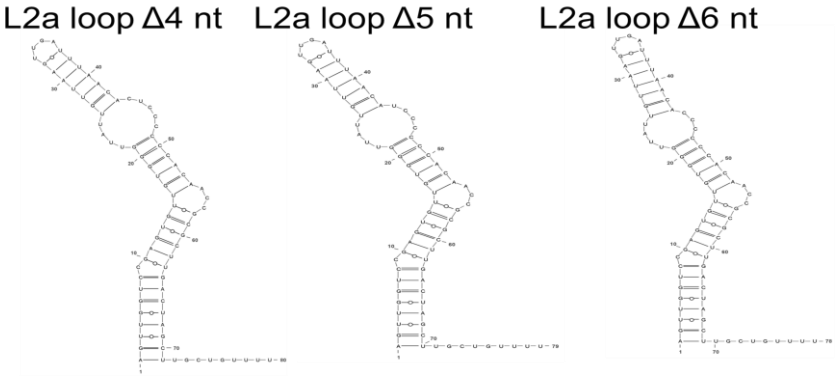
From the Northern blot results and relative quantification of L2a loop mutant derived YsRNAs compared to wild type hY5 RNA (**Figure 3.9**, **C and D**) it can be concluded that a bulge of one nucleotide might be sufficient for hY5 RNA cleavage. Intriguingly, the calculated ratio between full length hY5 RNA mutant levels compared to the YsRNA levels of the L2a loop mutants clearly showed that the mutants L2a loop Δ 4 nt, L2a loop Δ 5 nt and L2a loop Δ 6 nt were processed more efficiently than wild type hY5 RNA, but all the L2a loop mutants got processed except the L2a loop Δ 9 nt in which the loop was completely removed. Therefore, a bulge of at least one nucleotide is crucial for Y RNA cleavage.

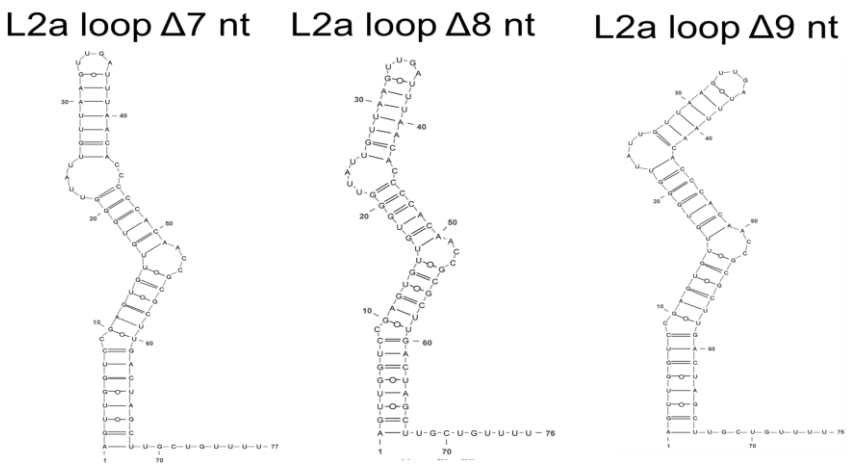
A


Internal loop L2a

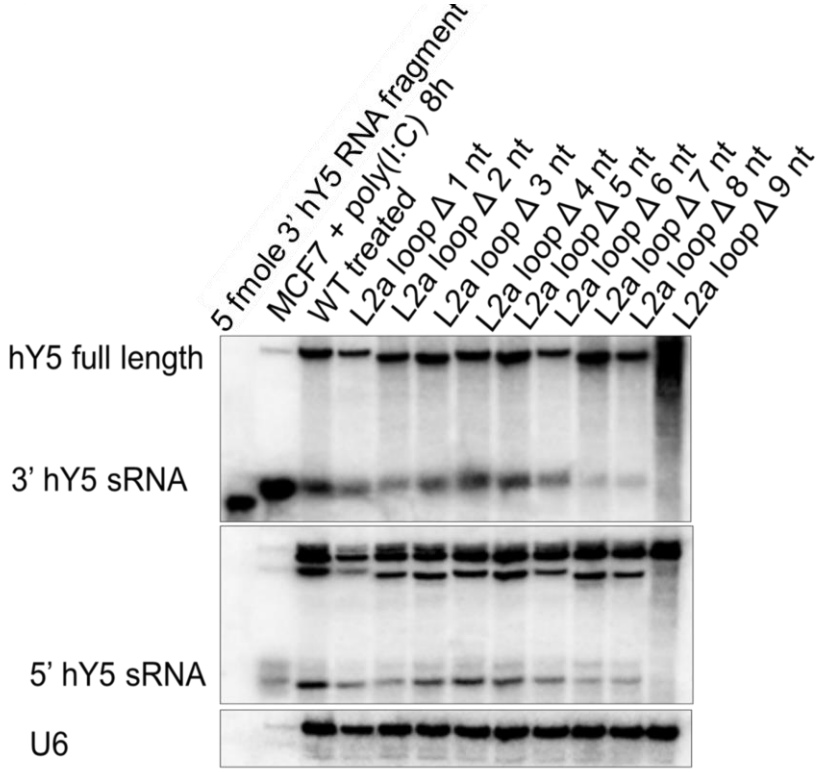
B


L2a loop Δ1 nt
L2a loop Δ 2 nt
L2a loop Δ 3 nt

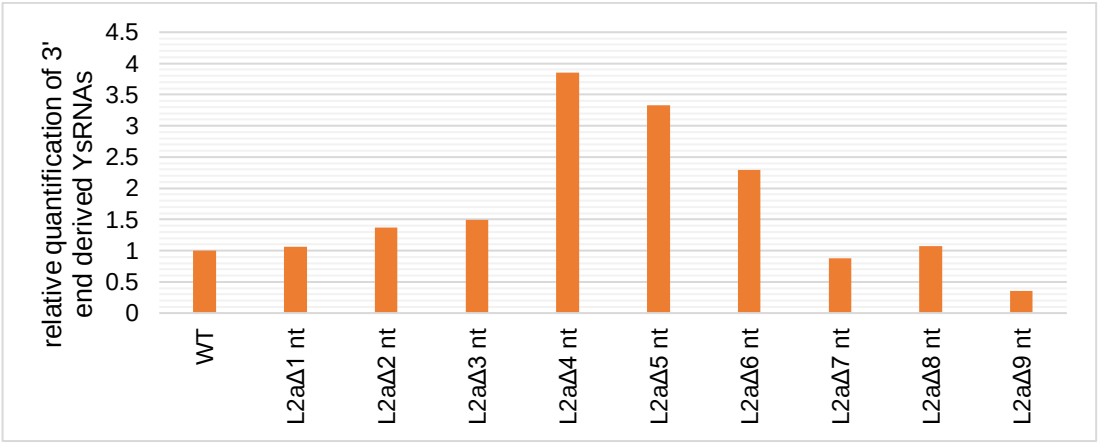

L2a loop Δ4 nt
L2a loop Δ5 nt
L2a loop Δ6 nt


L2a loop Δ7 nt
L2a loop Δ8 nt
L2a loop Δ9 nt

C


5 fmole 3' hY5 RNA fragment
MCF7 + poly(I:C) 8h
WT treated
L2a loop Δ 1 nt
L2a loop Δ 2 nt
L2a loop Δ 3 nt
L2a loop Δ 4 nt
L2a loop Δ 5 nt
L2a loop Δ 6 nt
L2a loop Δ 7 nt
L2a loop Δ 8 nt
L2a loop Δ 9 nt
hY5 full length
3' hY5 sRNA
5' hY5 sRNA
U6

D


relative quantification of 3' end derived YsRNAs
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
WT
L2aΔ1 nt
L2aΔ2 nt
L2aΔ3 nt
L2aΔ4 nt
L2aΔ5 nt
L2aΔ6 nt
L2aΔ7 nt
L2aΔ8 nt
L2aΔ9 nt

E

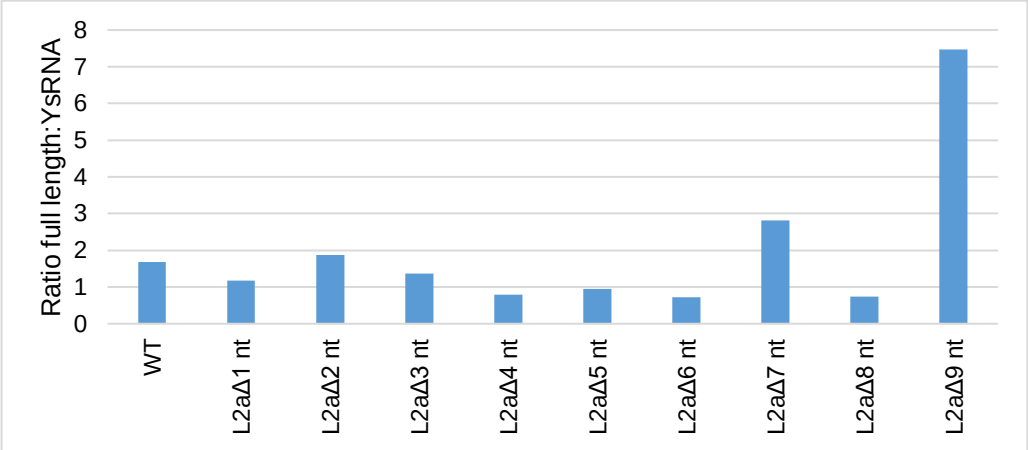


Figure 3.9. A bulge of one nucleotide is required for hY5 RNA cleavage.

A. Secondary structure of hY5 RNA.

The secondary structure of hY5 RNA was predicted with RNAfold and visualized using Varna. In this experiment the internal loop L2a indicated in red was mutated.

B. Predicted secondary structures of L2a loop deletion mutants.

The secondary structures of the L2a loop mutants Δ1-9 were predicted using RNAfold. The secondary structures were visualized using Varna.

C. Northern blot analysis of L2a loop deletion mutants.

As a size marker total RNA extracted from human MCF7 cells treated with poly (I:C) and a synthetic 3' end derived Y RNA fragment of 31 nt were used. The L2a loop deletion mutants and the wild type hY5 RNA plasmid were transfected into mouse cells and treated with staurosporine to induce YsRNA production. The total RNA was extracted and subjected to Northern blot. The Northern blot was probed with the probe complementary to the 3' and 5' end of the hY5 RNA. U6 was used in order to check for equal loading. The amount of YsRNAs derived from the 3' and 5' end of L2a loop Δ 1 nt to L2a loop Δ 8 nt mutants was less compared to the wild type hY5 RNA. For the mutant L2a loop Δ 9 nt no full length hY5 could not be detected.

D. Relative quantification of YsRNAs derived from the 3' end of L2a loop hY5 RNA mutants.

The Northern blot of 3' end derived L2a loop hY5 RNA YsRNAs mutants was quantified using the ImageQuant software. The optical density of each mutant was normalized to the loading control U6 and the wild type hY5 RNA derived YsRNAs. The normalized optical density of the wild type derived YsRNAs was set at 1.

E. Calculated ratio between the expression level of full length hY5 RNA mutants compared to the expression level of 3' end derived L2a loop hY5 RNA mutants.

The ratio of the normalized optical density of full length L2a loop hY5 RNA mutants and the derived YsRNAs was calculated and plotted on the graph.

3.2.6 The sequence of the L2a loop does not seem to be important for human hY5 RNA cleavage

The aim of this experiment was to investigate if the sequence at the hY5 RNA cleavage is important for cleavage. It was previously shown in sequencing experiments of MCF7 cells treated with poly (I:C) that the hY5 RNA cleavage from the 3' end mainly occurs between C50 and C51 generating YsRNAs with a length of 32-34 nt.

3.2.6.1 Mutagenesis of cytosines at the positions 50, 51 and 52 of hY5 RNA

I designed and generated several hY5 RNA mutants in which the cytosines at positions 50, 51 and 52 were substituted or deleted in order to test if cytosines at these positions are critical for Y RNA cleavage.

Another hypothesis to be examined was that the cytosines could be methylated at the cleavage sites and a demethylase could be in involved in Y RNA cleavage as it was previously shown for vault RNAs. In this study it was demonstrated that the RNA methyltransferase NSun2 is involved in the m^5C methylation of vault RNAs. In NSun2 deficient cells the processing of vault RNAs into svRNAs was inhibited suggesting that the methylation of vtRNAs at the cytosines is important for the generation of svRNAs (Hussain et al., 2013).

Most of the generated hY5 substitution mutants except M9 and M14 folded in a similar way than the wild type hY5 RNA (**Figure 3.10***,* **A**). The secondary structure analysis (**Figure 3.10**, **B**), Northern blot analysis and relative quantification (**Figure 3.10**, **C, D and E**) showed that the nucleotide sequence at positions 50, 51 and 52 did not seem to have any effect on hY5 RNA cleavage. All the mutants except the mutant M9 ΔLoopL2a were cleaved and generated YsRNAs (**Figure 3.10**, **C**). Only the mutants M5 and M7 were less efficiently processed to YsRNAs compared to the wild type hY5 (**Figure 3.10**, **C and D**). This could be explained that the introduced hY5 RNA substitution mutations changed the secondary structure of these mutants in such a way that the RNases that are involved in Y RNA cleavage are recruited less efficiently and cannot get access or the mutants M5 and M7 are less stable at the full length transcript level than the other hY5 RNA substitution mutants. This might be the case because the Northern blot already indicated that the full length levels of mutants M5 and M7 are much less compared to the wild type hY5 RNA full length transcript (**Figure 3.9**, **B**).

Therefore, the results indicated that the cytosine residues at the positions 50, 51 and 52 are not necessary for hY5 RNA cleavage from the 3' end. However, it has to be noted that the hY5 mutants M5 and M7 in which the C50 and C51 were mutated to G50 and A51 resulted in less full-length hY5 transcript levels and processed less efficiently compared to the other hY5 RNA mutants (**Figure 3.10**, **B, C and D**).

The secondary structure predictions I could show that when cytosines at positions 50, 51 or 52 are substituted or deleted the hY5 RNA mutants still maintain the L2a loop like the wild type hY5 RNA (**Figure 3.10**, **A and B**).

However, in the M9 ΔLoopL2a mutant that was generated by making it complementary to the sequence of loop L2b no full-length hY5 RNA transcript could be detected by Northern blot and relative quantification (**Figure 3.10**, **B, C and D**). The mutation in the M9 mutant changed the secondary structure in such a manner that it became a long double stranded RNA that could be further targeted for degradation.

The mutants that involved point mutations of C51 to either G, T or A (M1, M3, M7) did not have any effect on the hY5 RNA full length or processing from the 3' or 5' end of the hY5 RNA. Except the mutant M7 had slightly decreased full length hY5 mutant transcript levels and M7 hY5 mutant Y RNA sequences were less efficiently cleaved **(Figure 3.10, C and D).**

When C50 was mutated to A, T or G (M5, M6, M8) no effect on full-length hY5 transcript level or Y RNA processing was seen in the mutants M6 and M8 compared to the wild type. The mutant M5 showed less full length hY5 RNA mutant transcript levels and only a small amount of YsRNAs was generated **(Figure 3.10, C and D).**

In the mutants M11 and M12 the positions C50 and C51 of the hY5 RNA gene were mutated to G50 and G51 or G50 and T51. The Northern blot analysis also showed that the nucleotide sequence did not affect Y RNA cleavage from either the 3' end or the 5' end.

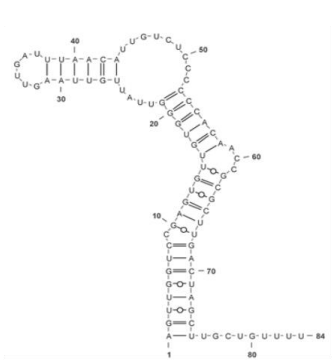
Interestingly, YsRNAs were still generated at the same amount compared to the wild type when two cytosines at position C50 and C51 close to the 3' end hY5 cleavage site in the mutant M13 were deleted. The predicted secondary structure of M13 folds very similarly to the hY5 wild type structure and contains 8 nt in the L2a loop whereas the wild type consists of 9 nt (**Figure 3.10**, **A and B**).

However, the deletion of three cytosines at the positions C50, C51 and C52 in the mutant M14 altered the secondary structure in a way that the single stranded L2a loop gets shortened from 9 nt to 3 nt which resulted in a lower accumulation of YsRNAs compared to the wild type (**Figure 3.10, B, C and D**). These results coincide with the previous Northern blot analyses of the L2a loop deletion mutants. Shortening of the L2a loop resulted in a decrease of YsRNA generation compared to the wild type and Y RNA cleavage still occurred when there was at least a bulge of 1 nt.

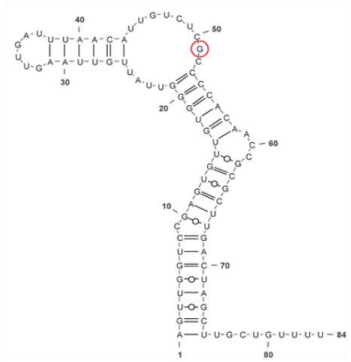
In conclusion, the mutagenesis of the cytosines indicated that these nucleotides at the position 50, 51 and 52 are not essential for hY5 RNA cleavage. Additionally, this might also suggest that m^5C methylation by the RNA methyltransferase NSun2 that was shown to be involved in vault RNA processing might not play a role in YsRNA generation.

A

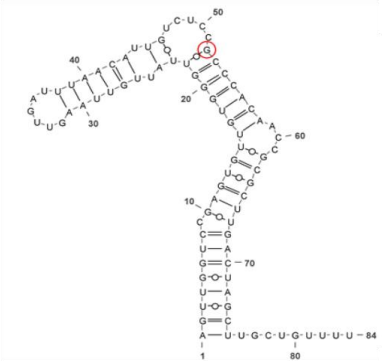
WT hY5 RNA



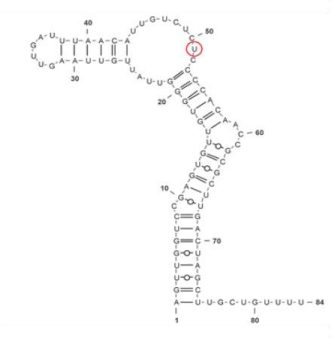
M1 C51--> G51



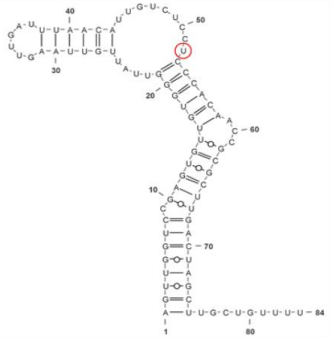
M2 C52-->G52



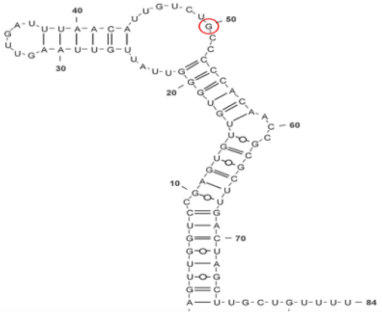
M3 C51-->U51



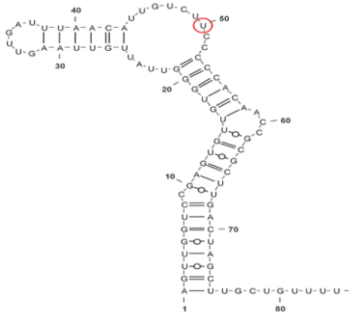
M4 C52-->U52



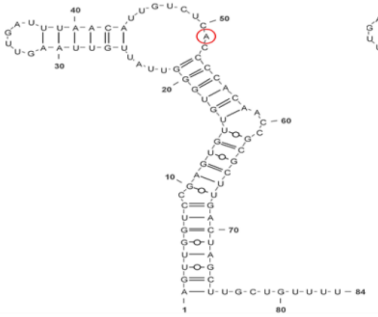
M5 C50-->G50



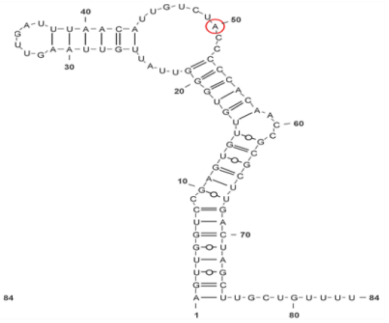
M6 C50-->U50



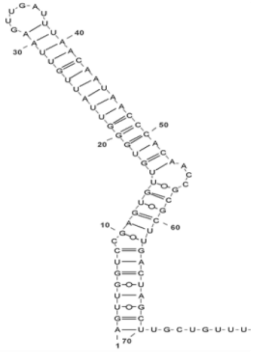
M7 C51-->A51



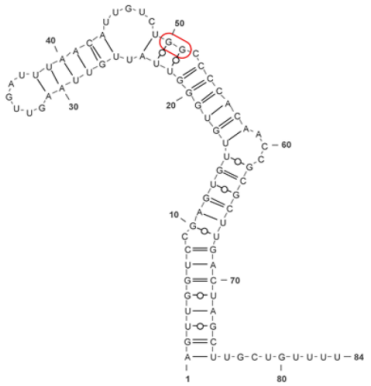
M8 C50-->A50



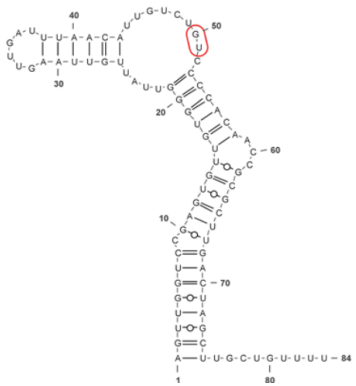
M9 Δloop2a



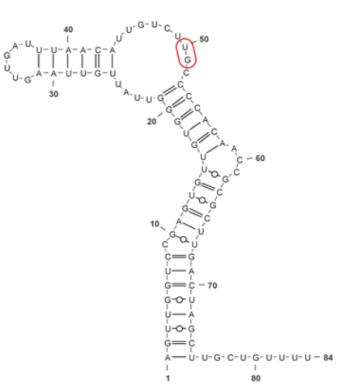
M10 C50C51--> G50G51



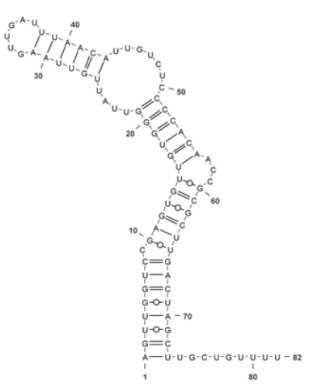
M11 C50C51--> G50U51



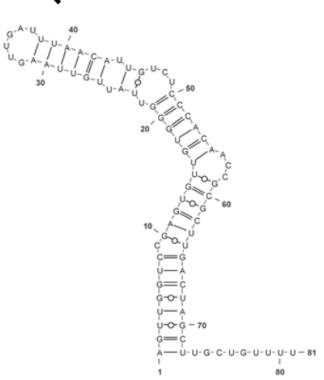
M12 C50C51--> U50G51



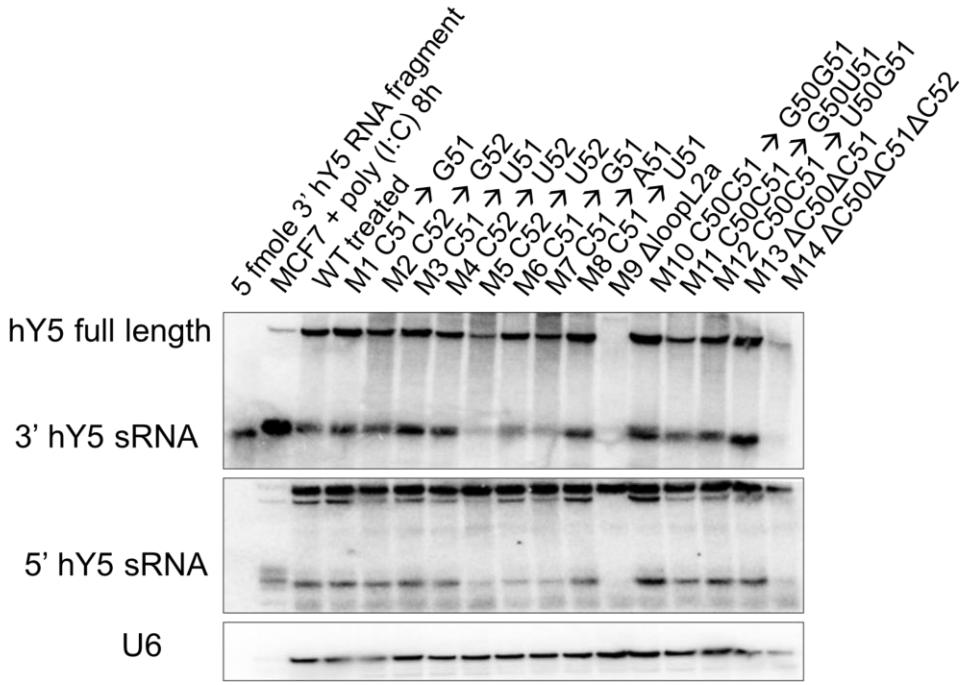
M13 ΔC50ΔC51



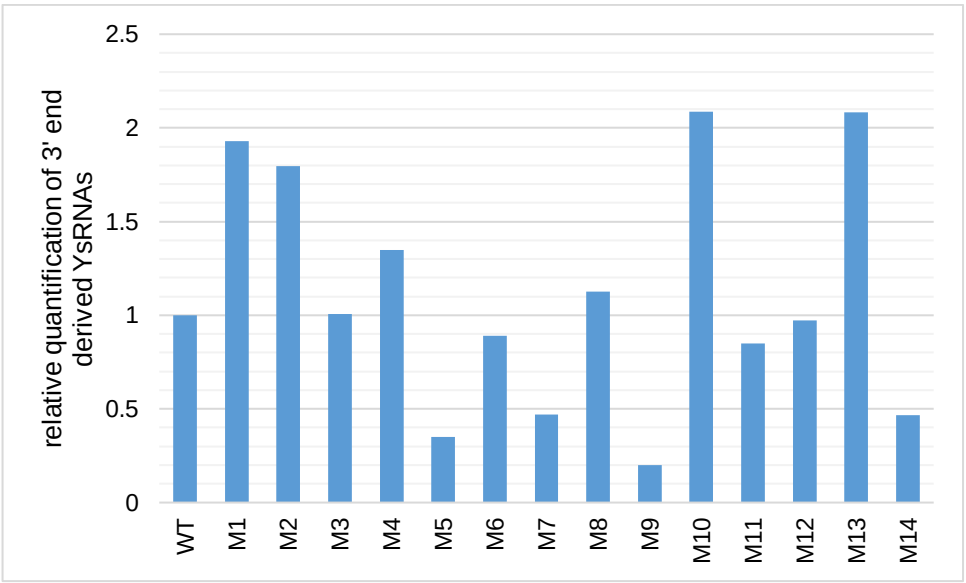
M14 ΔC50ΔC51ΔC52



B


5 fmole 3' hY5 RNA fragment
MCF7 + poly (I:C) 8h
WT treated
M1 C51 → G51
M2 C52 → G52
M3 C51 → U51
M4 C52 → U52
M5 C52 → U52
M6 C51 → G51
M7 C51 → A51
M8 C51 → U51
M9 ΔloopL2a
M10 C50C51 → G50G51
M11 C50C51 → G50U51
M12 C50C51 → U50G51
M13 ΔC50ΔC51
M14 ΔC50ΔC51ΔC52
hY5 full length
3' hY5 sRNA
5' hY5 sRNA
U6

C


relative quantification of 3' end derived YsRNAs
2.5
2
1.5
1
0.5
0
WT
M1
M2
M3
M4
M5
M6
M7
M8
M9
M10
M11
M12
M13
M14

D

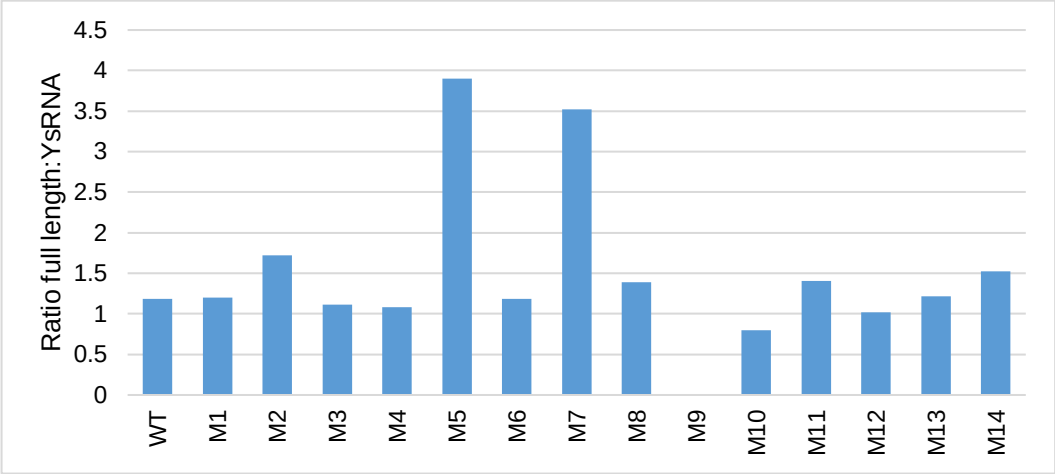


Figure 3.10. The cytosines at the positions 50, 51 and 52 are not essential for hY5 RNA cleavage.

A. Predicted secondary structures of L2a loop mutants.

The secondary structures of the L2a loop mutants were predicted using RNAfold and visualized using Varna. The mutated nucleotides are indicted with a red circle.

B. Northern blot analysis of L2a loop mutants.

The L2a loop mutant sequences were transfected into mouse cells and treated with staurosporine to induce Y RNA cleavage. Total RNA was extracted and subjected to Northern blot analysis to investigate the amount of YsRNAs that were derived from L2a loop mutants. The Northern blot was probed with the 3' end hY5 probe, 5' end hY5 probe and U6 as a loading control. As a positive control and size marker RNA extracted from MCF7 treated with poly (I:C) and a synthetic 3' hY5 derived fragment were included.

C. Relative quantification of YsRNAs derived from the 3' end of L2a loop hY5 RNA mutants.

The Northern blot of 3' end derived L2a loop hY5 RNA YsRNAs mutants was quantified using the ImageQuant software. The optical density of each mutant was normalized to the loading control U6 and the wild type hY5 RNA derived YsRNAs. The normalized optical density of the wild type derived YsRNAs was set at 1.

D. Calculated ratio between the expression level of full length hY5 RNA mutants compared to the expression level of 3' end derived L2a loop hY5 RNA mutants.

The ratio of the normalized optical density of full length L2a loop hY5 RNA mutants and the derived YsRNAs was calculated and plotted on the graph.

3.2.6.2 Mutagenesis of uridine at position 49 of hY5 RNA

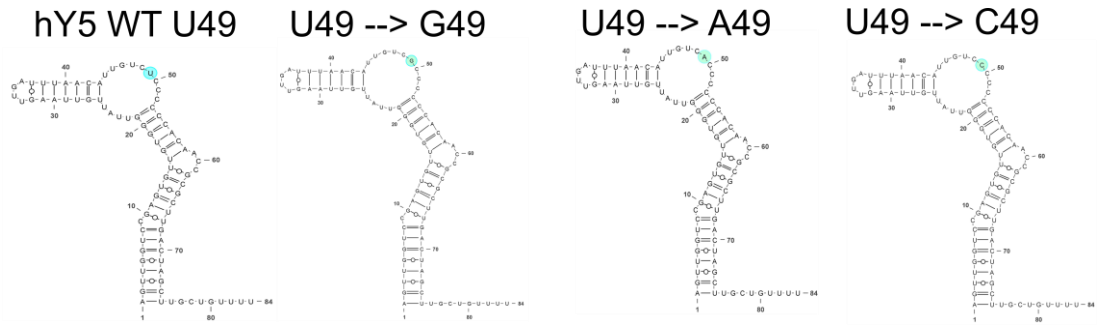
Recently, Donovan *et al*. reported that the ribonuclease RNase L is responsible for tRNA cleavage and Y RNA cleavage from the 5' end of hY5 RNAs. Since RNase L is known to cleave single-stranded RNAs after the UN^N in a sequence-specific manner it was investigated if hY5 RNA still generate YsRNAs when the uridine at position 49 is mutated to adenine, guanine or cytosine. Additionally, this UN^N motif was in proximity to the hY5 RNA cleavage position at C50 and C51 raising the question if RNase L is responsible for Y RNA cleavage from the 3' end. I generated hY5 RNA substitution mutants in which the uridine at position 49 was mutated to A, G and C and transfected them into 3T3 mouse cells. In a parallel experimental set up I transfected these hY5 mutants into wild type and RNase L $^{-/-}$ mouse embryonic fibroblast cells. The YsRNA production was tested by Northern blot analysis. Additionally, I analysed the secondary structures of these mutants using RNAfold.

The secondary structure analysis (**Figure 3.11**, **A**) revealed that the introduced mutations that position 49 did not change the secondary structure and all hY5 RNA mutants folded in the same way than wild type hY5 RNA.

Intriguingly, the Northern blot analysis of hY5 RNA mutants transfected into 3T3 mouse cells (**Figure 3.11**, **B**) showed that mutating the uridine at position 49 of the hY5 RNA to guanine, cytosine and adenine had no effect on 3' and 5' end Y RNA cleavage. This suggested that RNase L might not be critical for Y RNA cleavage from the 3' end and the UN^N sequence motif does not seem to be important.

From the Northern blot analysis of hY5 substitution mutants it can be concluded that uridine at position 49 can be replaced by adenine, guanine and cytosine without any effect on Y RNA expression and Y RNA cleavage from the 3' and 5' end. It was clearly shown that the hY5 substitution mutants were cleaved into Y RNA fragments from the 3' and 5' end.

A



B

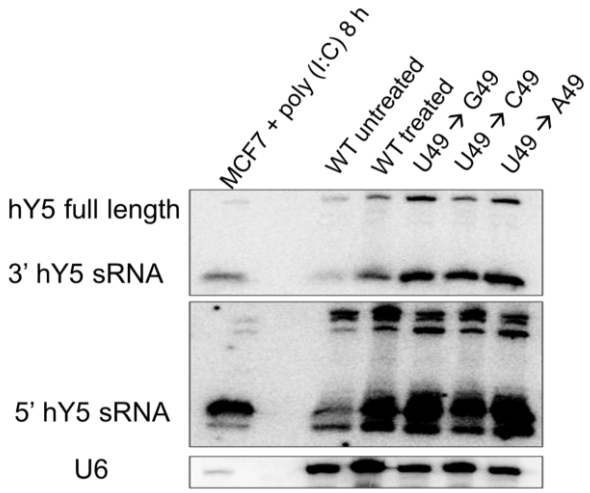


Figure 3.11. Substitution mutations of hY5 RNA at position 49 did not have any effect on hY5 RNA cleavage.

A. Predicted secondary structures of hY5 RNA substitution mutants at position 49.

The secondary structures were predicted with RNAfold and visualized using Varna. The mutated nucleotide is indicated with a turquoise circle.

B. Northern blot analysis of hY5 substitution mutants at position 49 in 3T3 mouse cells.

The hY5 mutants were transfected into 3T3 mouse cells, treated with staurosporine, total RNA was extracted after 8 h and subjected to Northern blot analysis. The Northern blot membrane was hybridized with the 3' end hY5 probe, 5' end hY5 probe and U6 as a loading control. As a size marker total RNA extracted from human MCF7 cells treated with poly (I:C) was used.

3.3 Discussion

From previous deletion and substitution mutagenesis analysis on the human Y5 RNA conducted by Carly Turnbull, a previous PhD student it was hypothesized that Y RNA cleavage correlates with secondary structure and is rather sequence independent (Turnbull, 2014). In several mutagenesis experiments it was observed that the stem-loop structure and Ro60 binding site were required for the generation of YsRNAs (Turnbull, 2014).

However, in order to establish sequence and structural requirements for human Y5 RNA cleavage and to identify proteins involved in Y RNA cleavage, the approach in generating individual hY5 mutants and testing their expression level by Northern blot analysis was quite time consuming, low throughput and only a few hY5 mutants could be tested at a time.

In recent years next generation sequencing of small RNA libraries has become an important tool for profiling and discovery of novel sRNAs in different biological systems. In our laboratory a high throughput mutagenesis approach on the human Y5 RNA has been designed to apply the NGS technology that enabled to examine thousands of different hY5 RNA mutants within a particular region of the hY5 RNA at the same time. Using a high throughput mutagenesis approach on the hY5 RNA instead of testing individual hY5 RNA mutants has the advantage to look at thousands of different mutants at the same time. This approach would help to identify a particular region or structural element of the hY5 RNA and proteins involved and responsible for Y RNA cleavage.

After NGS of the 3' hY5 high throughput mutagenesis libraries and bioinformatics analysis the hypothesis was confirmed that the Y RNA cleavage site correlated with the secondary structure of the mutated hY5 RNAs. The size class distribution showed that hY5 library 2 and 3 mutants mainly generated wild type-sized YsRNAs. Interestingly, hY5 library 1 mutants produced YsRNA reads that were 1-2 nt longer compared to mutant pool 2 and 3.

I generated and cloned the hY5 RNA mutant sequences with the five most abundant and five least abundant YsRNAs of each library individually and tested and compared their expression levels as well as YsRNA generation to the wild type hY5 RNA by Northern blot analysis and relative quantification.

I could confirm that the library 2 and 3 hY5 RNA mutant sequences generated wild type-sized YsRNAs whereas library 1 hY5 RNA mutants generated longer YsRNAs compared to the wild type.

I performed secondary structure prediction of 30 library 1 hY5 RNA mutants with the most abundant YsRNAs and I could clearly show that the mutations introduced in region 1 resulted in a change of the secondary structure.

In further detailed analysis of the secondary structural features of library 1 hY5 RNA mutants with the five most abundant and least abundant YsRNA reads I observed that the secondary structure of the hY5 library 1 mutants that generated the most abundant YsRNAs all consisted of a pyrimidine-rich internal loop that was shortened by 1-2 nt compared to the wild type hY5 RNA. Interestingly, in all of those mutants the internal loop was followed by a stem which was 5-6 nt long and had a high G-C content. In contrast, most of the hY5 library 1 mutants that only generated a low number of YsRNAs only contained a stem of 3 nt. This might suggest that a stem consisting of 5-6 nt and an internal loop might be required as structural features for proteins/ ribonuclease(s) to recognize hY5 RNA and generate YsRNAs from the 3' end of hY5 RNAs.

From the secondary structure analysis of the hY5 library 1 mutants I concluded that by the introduced mutations in library 1 the internal L2a loop got shortened by 1-2 nt compared to the internal loop of the wild type hY5 RNA. Thus, in library 1 hY5 RNA mutants the Y RNA cleavage occurs 2-3 nt more upstream compared to the wild type hY5 RNA but still two nucleotides above the stem. This would suggest that protein(s)/ribonuclease(s) involved in hY5 RNA cleavage from the 3' end might recognize the stem and cleave hY5 RNAs between the 2^{nd}/3^{rd} nt above the stem.

The results from the next generation sequencing, secondary structure prediction, Northern blot analyses confirmed that for hY5 RNA mutants in library 2 and 3 the Y RNA cleavage mainly occurred at the position C50/C51.

By Northern blot analyses of the library 1 hY5 RNA mutants I demonstrated that Y RNA cleavage occurs between 2^{nd}/ 3^{rd} nt above the stem of 5-6 nucleotides and showed that the Y RNA cleavage correlates with the secondary structure.

In addition, by relative quantification of Northern blots comparing full length library 1 hY5 RNA mutant levels to YsRNAs compared to the wild type hY5 RNA I could show that library 1 hY5 RNA mutants with the most abundant YsRNAs were as efficiently processed as wild type hY5 RNA. This further suggested that 3' end hY5 RNA cleavage correlates with secondary structure.

Interestingly, hY5 library 1 mutations that resulted in the same secondary structure as the wild type hY5 sequence showed wild type sized YsRNAs whereas hY5 library 1 mutations that have different secondary structure from the wild type hY5 RNA generated longer hY5 RNA fragments. Thus, it further nicely confirmed the hypothesis that hY5 RNA cleavage from the 3' end correlated with the secondary structure.

The secondary structure predictions suggested that the nucleotide sequence introduced in mutant pool 1, 2 and 3 did not have any effect on Y RNA cleavage as long as the secondary structure contained a five nucleotide stem region preceded by an internal loop. These results indicate that proteins involved in Y RNA cleavage might recognize the stem and cleave the hY5 RNA two to three nucleotides in the L2a loop.

In this part of the work I clearly demonstrated that the secondary structure of the hY5 RNA correlates with 3' end hY5 RNA cleavage and I was able to identify and establish secondary structure features and elements.

The next steps would involve the investigation of the proteins/endoribonuclease(s) involved in 3' end hY5 RNA cleavage that recognize the stem region and cleave in a single-stranded loop region using RNA immunoprecipitation experiments combined with mass spectrometry. One approach to look into Y RNA interaction partners of Y RNAs and potential ribonucleases in Y RNA cleavage would be to transfect biotinylated Y RNAs and carry out staurosporine treatment. The next step would involve immunoprecipitation experiments using streptavidin beads of untreated and treated samples followed by mass spectrometry.

Further experiments would involve investigations into the function of 3' end derived hY5 RNA fragments because until now in the literature no function of 3' end derived hY5 fragments has been described. In contrast, only the 5' end derived hY5 RNA fragment was reported to trigger apoptosis (Chakrabortty et al., 2015).

Additionally, 5' end derived Y RNA fragments have been shown to be highly enriched in extracellular vesicles of several cancers (Dhahbi et al., 2013) and diseases (Cambier et al., 2017) and have been observed to be involved in immunity (Hizir et al., 2017).

In order to further investigate the sequence and structural requirements for 3' end hY5 RNA cleavage I carried out additional mutagenesis experiments to explore if there was a minimal nucleotide requirement for Y RNA cleavage to occur.

The loop 2a was shortened by deletion mutagenesis and hY5 RNA mutants were subjected to Northern blot analyses. The obtained sRNA Northern blot results and their relative quantifications by densitometry suggested that the removal of 8 nt from the L2a loop in general only resulted in a small decrease of YsRNA production from the 3' and the 5' end of the hY5 RNA. However, the removal of 9 nt had a dramatic effect on YsRNA generation from the 3' and 5' end of the human Y5 RNA and not even the full length hY5 RNA was stable anymore. This phenomenon might be due to the fact that the Loop L2a Δ9 nt mutant folds similar to long double stranded RNA such as viral double stranded RNA that might be processed by RNase L for degradation.

From the L2a loop deletion mutagenesis I concluded that a bulge of one nucleotide is required for Y RNA cleavage or to have a stable Y RNA that can be processed.

In parallel to the L2a loop deletion experiments I studied in further experiments, if the nucleotide sequence at the positions 50, 51 and 52 of the L2a loop which are in proximity to the 3' end hY5 RNA cleavage site were important for 3' end hY5 RNA cleavage. It was shown in previous experiments that the hY5 RNA cleavage from the 3' end mainly occurs between C50 and C51 producing YsRNAs of 33-34 nt.

In the Northern blot analysis and relative quantification of 3' end derived YsRNAs I observed that the nucleotide sequence at the positions 50, 51 and 52 did not seem to have any effect on hY5 RNA cleavage. Therefore, this suggested that cytosine residues at the positions 50, 51 and 52 are not necessary for hY5 RNA cleavage from the 3' end. Thus, we could conclude that m^5C methylation of cytosines that is mediated by the RNA methyltransferase NSun2 and generation of vault RNA derived small RNAs might not be involved in YsRNA generation.

Intriguingly, when all three cytosines at the positions C50, C51 and C52 were deleted a small decrease of YsRNA accumulation compared to wild type was observed, but Y RNA cleavage from the 3' and 5' end of hY5 RNAs still occurred. This result was consistent with the previous Northern blot analyses obtained from the loop L2a deletion mutagenesis experiments that revealed that a bulge of 1 nt was sufficient for Y RNA cleavage to occur.

Interestingly, when the loop L2a was completely deleted in the mutant M9 by generating a sequence complementary to the sequence of the loop L2b no full length hY5 RNA transcript could be detected by Northern blot. The predicted secondary structures of this mutant indicate that the structure was altered dramatically compared to the wild type hY5 RNA and it formed a long double stranded RNA region that might be targeted for degradation resulting in the loss of Y RNA stability.

As in 2017 Donovan *et al*. claimed that RNase L is responsible for Y RNA cleavage from the 5' end (Donovan et al., 2017) I investigated if this ribonuclease is also involved in the generation of Y RNA fragments from the 3' end. RNase L is known to cleave single-stranded RNAs in a sequence-specific manner after UN^N sequence motif. The very same motif was found close to the 3' end Y RNA cleavage site.

In order to prove that RNase L is critical for the 3' end hY5 RNA cleavage I mutated the uridine at position 49 to adenine, guanine and cytosine and tested the generated hY5 substitutions mutants by Northern blot. The obtained Northern blot experiments clearly showed that YsRNA fragments were generated from both the 3' and 5' end of hY5 RNA substitution mutants. These experiments suggested that uridine at position 49 is not essential for Y RNA cleavage from the 3' end of hY5 RNA. Thus, RNase L might not be responsible for Y RNA cleavage from the 3' end of hY5 RNAs and other proteins/ribonucleases are involved in Y RNA cleavage from the 3' end of hY5 RNA.

Chapter 4 Establishing sequence and structural requirements for 5' end human Y5 RNA cleavage

4.1 Introduction

4.1.1 Extracellular vesicles

Extracellular vesicles (EVs) are a heterogeneous population of membrane-enclosed vesicles comprised of a phospholipid bilayer. They are characterized to release cargoes such as nucleic acids, lipids or proteins from the plasma membrane of donor cells to neighbouring or distinct recipient cells. EVs have been identified in both eukaryotes and prokaryotes and the secretion of EVs is evolutionarily conserved (Gould and Raposo, 2013). EVs are derived from many different cell types and have been isolated from body fluids such as urine, blood, saliva, semen or breast milk (Colombo et al., 2014).

According to their size, origin, morphology, purification, function and cargo EVs can be roughly classified into exosomes, microvesicles and apoptotic bodies. However the classification of EVs is still controversial in the field and there are still limitations in the purification of EVs (Gould and Raposo, 2013).

Exosomes (30-100 nm in diameter) are formed during endosomal pathways within large endosomal compartments called multivesicular endosomes or multivesicular bodies (MVBs). They are released into the extracellular space by fusion of MVBs with the plasma membrane (Johnstone et al., 1987). Microvesicles are larger than exosomes (100-1000 nm) and are released by budding or fission of the plasma membrane. Apoptotic bodies are generated during apoptosis and vary in size between 1000 nm and 5000 nm (Gould and Raposo, 2013; Colombo et al., 2014).

EVs are key mediators of intercellular communication and are involved in multiple physiological and pathological conditions. EVs play an important role in development, cell homeostasis, neuronal signalling and immunity (Yanez-Mo et al., 2015; Maas et al., 2017).

However, EVs have also been associated with inflammatory diseases (Buzas et al., 2014), neurodegenerative diseases (Coleman and Hill, 2015) and cancer (Minciacchi et al., 2015).

In the last years it was shown that cancer-cell derived EVs have been implicated in the tumour microenvironment and promote tumour progression, angiogenesis, cell invasion and metastasis (Tkach and Théry, 2016).

The recent development of next generation sequencing technologies in the last decade allowed the identification of a wide range of low abundant RNA species in EVs isolated from different types of body fluids. A breakthrough in the field of EVs was the discovery that exosome-derived mRNAs and miRNAs could be transferred between cells and the released mRNAs were translated into proteins in the recipient cells (Ratajczak et al., 2006; Valadi et al., 2007).

Due to the association of EVs with many different diseases including cancer, EV associated miRNAs have emerged as novel diagnostic biomarkers (Kinoshita et al., 2016) with a huge potential for therapeutic approaches (Fais et al., 2016).

Surprisingly, high throughput RNA sequencing of small RNAs derived from immune cell EVs revealed that miRNAs represented only a small fraction of the RNA repertoire in the EVs compared to cellular RNA. Unexpectedly, the EVs contained several other classes of regulatory non-coding RNAs, such as small nucleolar RNAs, small nuclear RNAs, transfer RNAs, ribosomal RNAs, long non-coding RNAs, circular RNAs, vault RNAs as well as Y RNAs (Nolte-'t Hoen et al., 2012; Kim Kyoung et al., 2017). Among them, vault RNAs and Y RNAs were found highly enriched in the immune cell derived EVs. In addition, Nolte-t' Hoen *et al*. also identified small RNA fragments derived from longer transcripts, such as Y RNA fragments, tRNA fragments or a vault RNA derived sRNA fragment enriched in immune cell EVs. Among them, Y RNA fragments were the most abundant small RNA fragments in EVs suggesting a functional role of these small RNAs derived from longer transcripts.

4.1.2 Functions of 5' end derived Y RNAs in extracellular vesicles

Vojtech *et al*. showed by next generation sequencing of exosomes derived from semen that 5' end derived Y RNA fragments were highly enriched (Vojtech et al., 2014). Interestingly, in human serum and plasma of healthy donors, Y RNA fragments derived from the 5' end of specific Y RNA genes and Y RNA pseudogenes with sizes of 27 nt and 30-33 nt were identified. Surprisingly, in the plasma these 5' end derived Y RNA fragments were not incorporated within EVs but circulated in the blood as part of a large complex with a mass of 100-130 kDa.

In the same study it was shown that 5' end derived Y RNA fragments were more abundant in human serum compared to mouse serum while the opposite was observed for 5' end tRNA halves. It needs to be further investigated why the composition of circulating sRNAs was different between human and mouse sera (Dhahbi et al., 2013). Specific tRNA halves and Y RNA fragments were found to be associated with breast cancer and differentially expressed compared to healthy individuals (Dhahbi et al., 2014).

Dhahbi *et al*. observed that in human serum 5' end derived Y RNA fragments are about twenty times more abundant compared to 3' end derived Y RNA fragments (Dhahbi et al., 2013). Interestingly, RNA sequencing of five breast cancer patients compared to control samples revealed that mostly 3' end derived Y RNA fragments were upregulated in five diagnosed breast cancer patients whereas 5' end derived Y RNA fragments showed downregulation compared to healthy patient samples (Dhahbi et al., 2014). However, any function or mechanism or biogenesis pathway remained obscure.

A recent study investigated the processing and transfer of hY5 RNA fragments within EVs derived from primary and cancer cell lines. It was shown that hY5 RNAs were highly abundant and enriched in EVs. Shorter fragments (23, 29 and 31 nt) derived from the 5' and 3' end of hY5 were also detected in EVs (Chakrabortty et al., 2015). In the same study hY5 RNA fragmentation occurred in the EVs instead of inside the cells.

Notably, the 5' end derived hY5 RNA fragments contained a conserved 8 nt RNA sequence motif (5'-GUUGUGGG-3') that was previously reported to be also involved in the initiation of chromosomal replication (Gardiner et al., 2009). The same motif was shown to be important for the processing of full-length hY5 RNAs into smaller functional Y RNA fragments (Chakrabortty et al., 2015). Interestingly, the exposure of human healthy cells with EVs derived from cancer cells resulted in rapid cell death. However, cancer cells did not show any phenotype when they were treated with EVs from normal or cancer cells.

Intriguingly, it was shown that only a synthetic version of the 23 nt- and 31 nt 5' end derived hY5 fragment and not the 3' end derived hY5 fragment nor a double stranded version of 3' and 5' hY5 fragment triggered cell death in primary cells in a dose-dependent manner through an apoptotic pathway.

Further, it was observed that the conserved 8 nt motif in the 5' end derived Y RNA fragments was required for the induction of cell death (Chakrabortty et al., 2015).

The study conducted by Chakrabortty *et al*. indicated that there seems to be a correlation between 5' end derived Y RNA fragments and cancer, since 5' end derived hY5 RNA fragments only derived from cancer cell EVs trigger rapid cell death in normal cells. It was hypothesized that the 5' end derived hY5 RNA fragment induced cell death might help tumours to invade tissues and to establish a microenvironment (Chakrabortty et al., 2015).

In 2012 the Dalmay laboratory already showed that Y RNAs are also generated from the 5' end of human Y RNAs in several cancerous and non-cancerous cell lines as previously described (Nicolas et al., 2012). However, because of the higher abundance of 3' end derived hY5 fragments compared to 5' end derived hY5 RNA fragments the research focus of our laboratory was mainly on the 3' end derived hY5 RNA fragments.

Due to the fact that Chakrabortty *et al*., gave some evidence that 5' end derived hY5 RNA fragments are functional and require a conserved 8 nt motif in order to induce cell death, in this part of the work I aimed to investigate if there was any sequence and/or secondary structure requirement for hY5 RNA cleavage from the 5' end. In order to pursue this, I designed and performed a high throughput mutagenesis approach on the 5' end of the human Y5 RNA.

4.2 Results

4.2.1 Generation of 5' end hY5 mutant plasmid pools

I designed the 5' end hY5 high throughput mutagenesis approach in a similar way to the 3' end hY5 mutagenesis study. Previous next generation sequencing performed in our laboratory showed that MCF7 cells treated with poly (I:C) generated hY5 RNA fragments from the 3' and 5' end. The main cleavage product from the 5' end of the hY5 RNA was 32 nt.

Thus, I selected three different five nucleotide regions immediately upstream of the determined cleavage site for random mutagenesis (**Figure 4.1**, **A and B**). The five nucleotide substitution mutations were introduced using primers containing degenerate nucleotides at each of the five positions.

The PCR products including the promoter region and 30 bp downstream terminator region were amplified from the hY5 wild type expression construct by overlap extension PCR (**Figure 4.1**, **B**) in three replicates for each mutant pool.

In order to check if the five nucleotide substitution mutations were introduced correctly and to see if the degenerate nucleotides at each of the five positions are equally distributed, I sent the generated PCR products of each library replicate for Sanger sequencing and analyzed the sequences and chromatograms.

I cloned the A-tailed PCR products into pGEMT easy vector, transformed them into *E. coli*. Apart from the PCR products I also sent ten plasmid DNA clones of each library replicate for Sanger Sequencing and checked if the sequence and introduced mutations were correct.

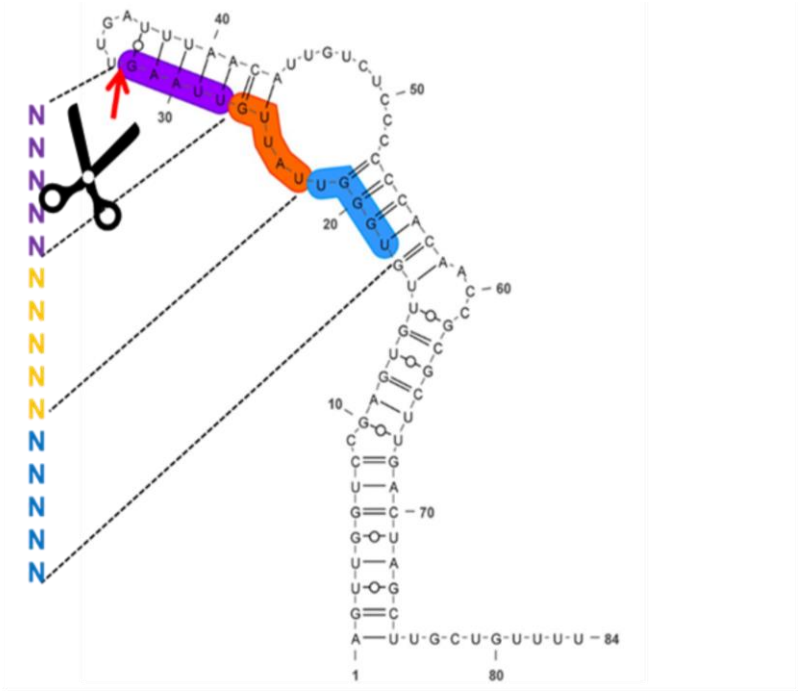
However, the analysis of the Sanger sequencing results of the mutant pool PCR products and purified plasmid DNA samples of each mutant pool replicate revealed that for mutant pool 4 and 5 the nucleotide composition at each of the five positions was significantly biased (**Figure 4.1**, **C and D**).

After literature search and contacting gene synthesis companies I was suggested to use "hand-mixed" primers containing degenerate nucleotides instead of custom synthesized primers in order to overcome the bias. The primers would be synthesized in a way that each nucleotide would be added individually by hand at a percentage of 25% at the five nucleotide mutagenesis region by hand and not randomized by the oligo synthesis machine.

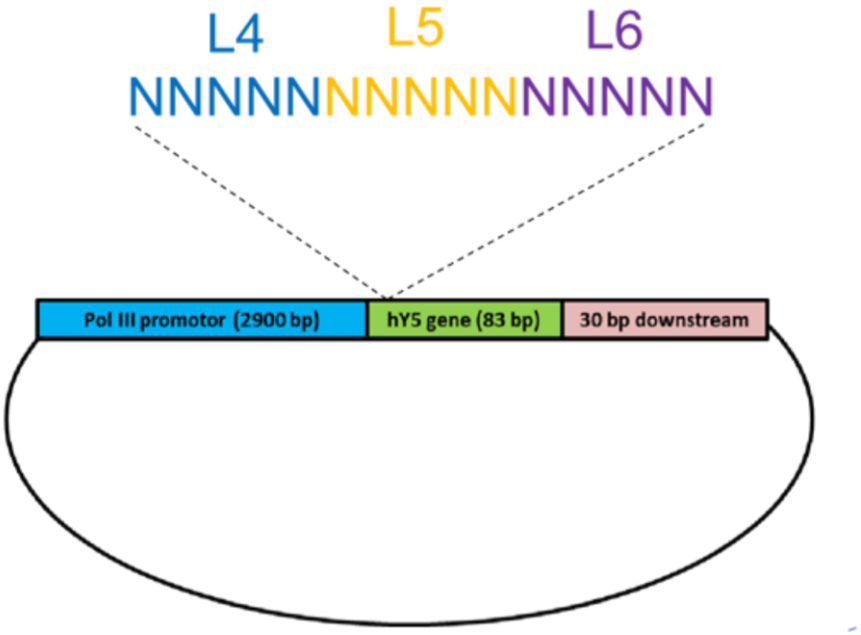
I repeated the PCR reactions for mutant pool 4 and 5 with newly designed "hand-mixed" primers and sent the PCR products of all mutant pool replicates for Sanger sequencing in order to investigate if the bias at the mutagenized regions could be improved.

The chromatograms of library 4, 5 and 6 mutant pool PCR products and Sanger Sequencing results of fifty isolated plasmid DNA clones of library 4, 5 and 6 using "hand-mixed" primers indicated that the distribution of the four possible nucleotides for library 4 and 5 was improved and more equally distributed (**Figure 4.1**, **E and F**) compared to the machine-mixed custom synthesized primers that were used for the PCR previously (**Figure 4.1**, **C and D**).

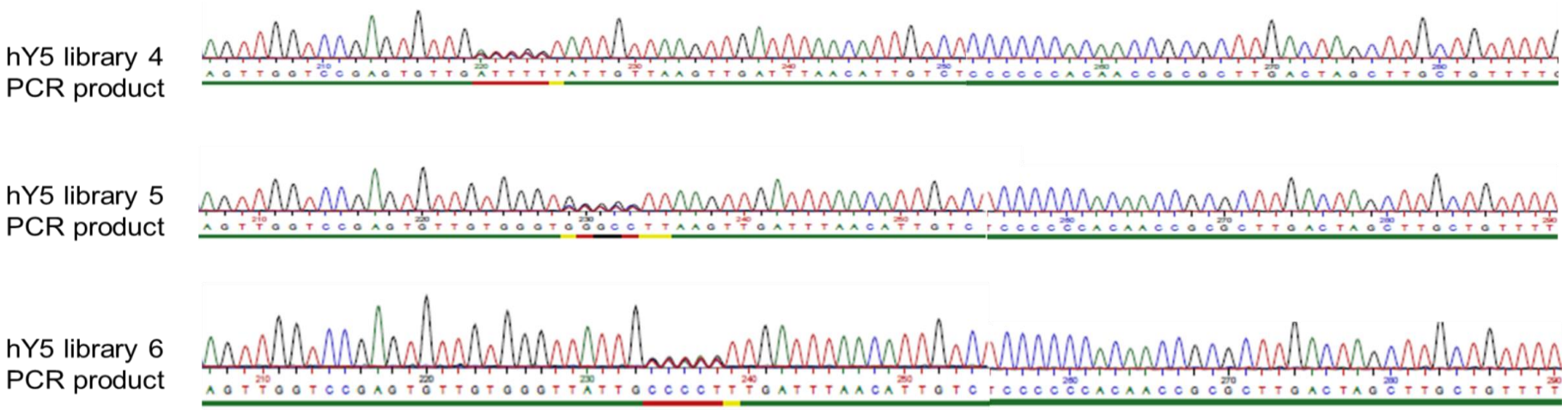
A


N
N
N
N
N
N
N
N
N
N
N
N
N
N
N
40
30
50
20
60
10
70
84
1
80

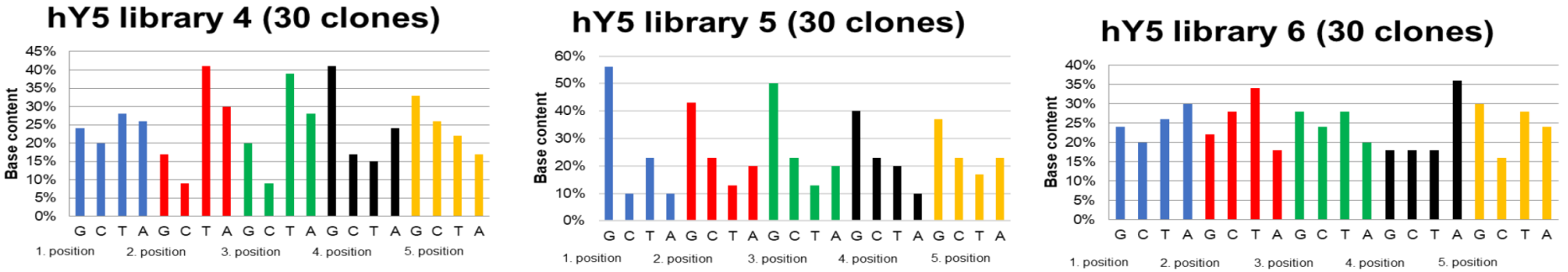
B


L4
L5
L6
NNNNNNNNNNNNNNN
Pol III promotor (2900 bp)
hY5 gene (83 bp)
30 bp downstream

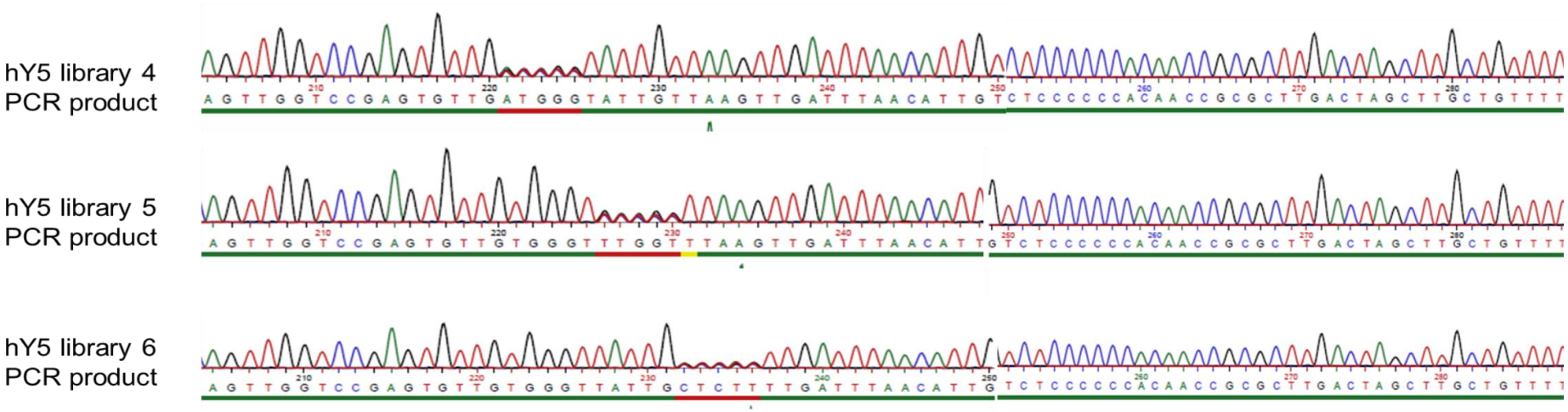
C


hY5 library 4
PCR product
hY5 library 5
PCR product
hY5 library 6
PCR product

D


hY5 library 4 (30 clones)
Base content
45%
40%
35%
30%
25%
20%
15%
10%
5%
0%
G C T A G C T A G C T A G C T A G C T A
1. position
2. position
3. position
4. position
5. position
hY5 library 5 (30 clones)
Base content
60%
50%
40%
30%
20%
10%
0%
G C T A G C T A G C T A G C T A G C T A
1. position
2. position
3. position
4. position
5. position
hY5 library 6 (30 clones)
Base content
40%
35%
30%
25%
20%
15%
10%
5%
0%
G C T A G C T A G C T A G C T A G C T A
1. position
2. position
3. position
4. position
5. position

E


hY5 library 4
PCR product
hY5 library 5
PCR product
hY5 library 6
PCR product

F

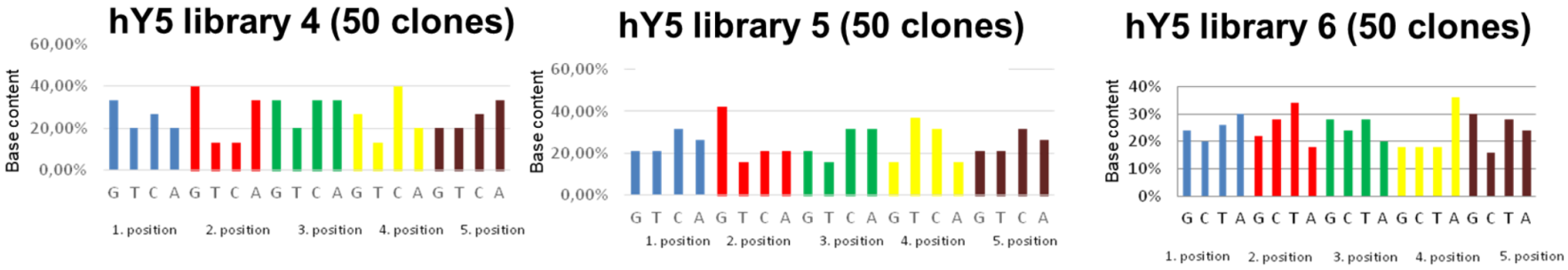

hY5 library 4 (50 clones)
hY5 library 5 (50 clones)
hY5 library 6 (50 clones)
Base content
1. position
2. position
3. position
4. position
5. position

Figure 4.1. Design and generation of the 5' end hY5 high throughput mutagenesis approach.

A. Predicted human hY5 RNA secondary structure with the regions that were mutated for the 5' end high throughput mutagenesis analysis.

In each of these regions of library 4 (blue), library 5 (orange) and library 6 (purple), five nucleotide substitution mutations were introduced resulting in 1024 (4^5) possible combinations for each library.

B. Schematic of the pGEM-hY5-WT vector containing the mutated regions of hY5 RNA library 4, 5 and 6.

The mutations for library 4 (blue), library 5 (orange) and library 6 (purple) were introduced by mutagenesis PCR using primers containing random nucleotides (N).

C. Chromatograms of generated PCR products from library 4, 5 and 6 using custom synthesized primers.

For library 4 and 5 it can be seen that the degenerate random nucleotides are not equally distributed and that some bases are overrepresented when machine-mixed degenerate primers were used for PCR, whereas in library 6 the degenerate nucleotides showed more or less equal proportions at each position.

D. Sanger sequencing results of hY5 library mutants generated from library 4, 5 and 6 using custom synthesized primers.

Ten plasmid DNA samples isolated from each library replicate of mutant pool 4, 5 and 6 were sent for Sanger Sequencing and analysed if the nucleotide composition at each of the five positions is equally distributed or biased. The graph shows the base composition on the x-axis as percentage of all the clones analysed. On the y-axis each position of the mutated five-nucleotide mutagenesis region is given.

E. Chromatograms of generated PCR products from library 4, 5 and 6 using hand-mixed synthesized primers.

When hand-mixed degenerate random primers were used for PCR, the proportions of each of the four bases at each of the positions of the five-nucleotide region for library 4 and 5 improved greatly and no overrepresentation of any nucleotide could be observed anymore.

F. Sanger sequencing results of hY5 library mutants generated from library 4, 5 and 6 using hand-mixed synthesized primers.

Fifty plasmid DNA samples isolated from one library replicate of mutant pool 4, 5 and 6 were sent for Sanger Sequencing and analysed if the nucleotide composition at each of the five positions is equally distributed or biased. The graph shows the base composition on the x-axis as percentage of all the clones analysed. On the y-axis each position of the mutated five-nucleotide mutagenesis region is given. The Sanger Sequencing results showed that hand-mixed random nucleotides for generating library 4, 5 and 6 mutant pool PCR products dramatically reduced the bias of nucleotide composition in each mutated position.

4.2.2 Generating 5' end hY5 mutant plasmid pools

Similar to the 3' end hY5 high throughput mutagenesis approach I generated 5' end hY5 plasmid libraries and full length hY5 cDNA libraries in order to check if all the 1024 possible hY5 mutant sequences were cloned successfully and to allow us to normalize the YsRNA datasets in a certain way and to account for occurring biases in the datasets.

For generation of the 5' end hY5 mutant plasmid pools I harvested around 10000 colonies for each library replicate in order to ensure that all the 1024 possible hY5 mutant combinations are present. Statistical analysis that was performed for the 3' end hY5 high throughput mutagenesis approach suggested that around 4850 positive colonies need to be harvested in order to be sure that all the 1024 possible sequences will be present with a probability of 99%.

I prepared the 5' end hY5 plasmid bias cDNA libraries as previously described and sent for next generation sequencing. The bioinformatics analysis of the plasmid bias libraries showed that all the 1024 different hY5 mutant sequences were present in the plasmid libraries that were generated from the 5' end of the hY5 RNA. However, the plasmid bias libraries were hugely biased for the hY5 wild type sequence in hY5 library pools 4 and 5. Among all the three replicates the hY5 wild type sequence was roughly six to seven times more abundant in hY5 mutant pool 4 and 5 than all the other 1023 hY5 mutant sequence combinations. For mutant pool 6 this prominent hY5 wild type sequence bias was not observed and all the 1024 hY5 mutant sequence combinations were present and statistically equally distributed.

Consequently, I had to modify the cloning strategy for mutant pool 4 and 5 of the 5' end high throughput mutagenesis approach in order to ensure that the wild type sequence is not overrepresented anymore.

The first cloning strategy was based on a plasmid containing the hY5 full length wild type sequence. The mutations for the 5' end mutagenesis approach were introduced by overlap extension PCR (**Figure 4.2, A**). Due to the fact that overlap extension PCR (**Figure 4.2, B**) in general is not very efficient any small traces of the full length hY5 wild type sequence will result in amplification of the hY5 wild type sequence.

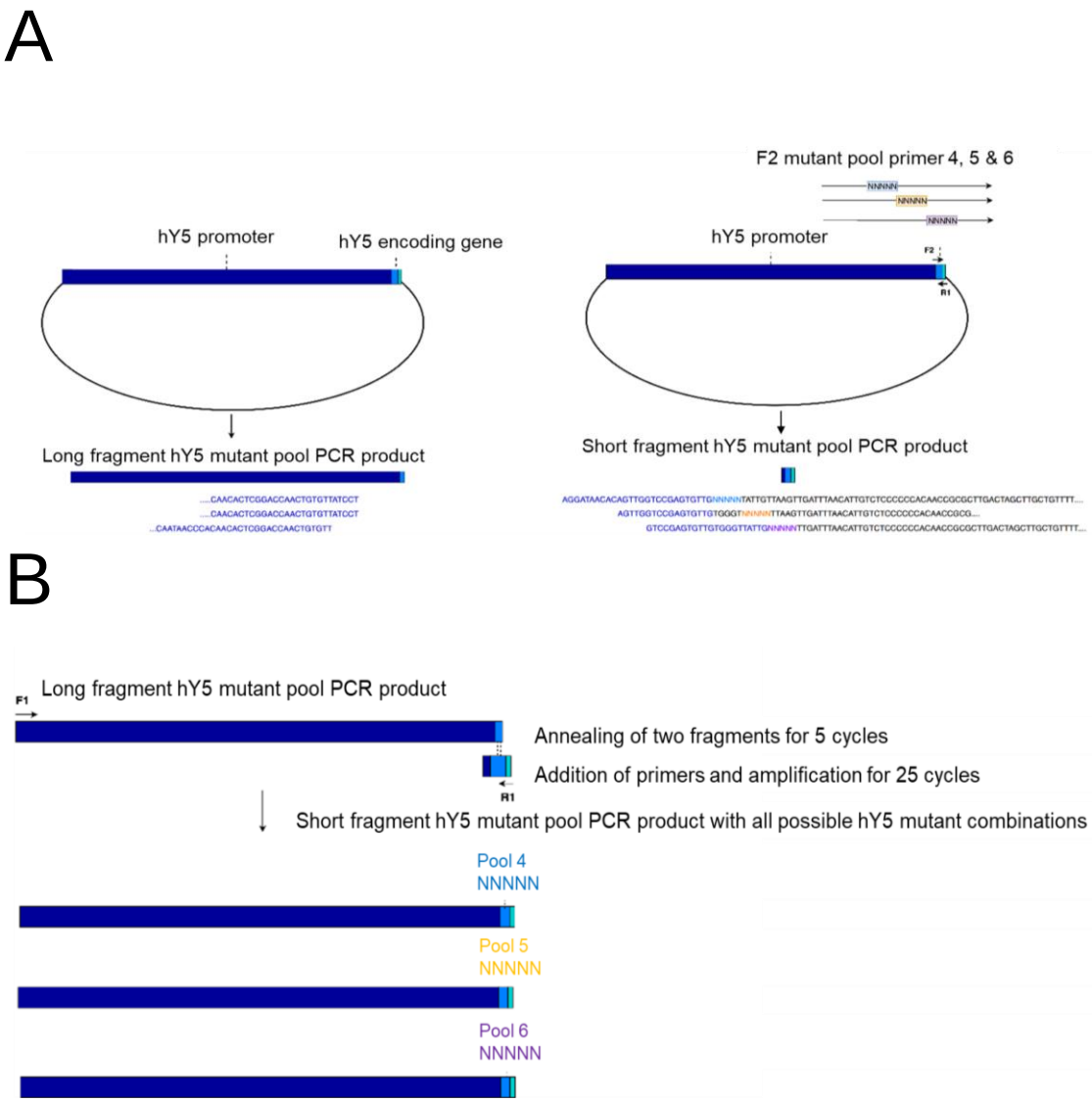


Figure 4.2. First cloning strategy for 5' end high throughput mutant pool generation.

A. Generation of the long fragment and short fragment PCR products for 5' end high throughput mutagenesis approach.

As a template for PCR the pGEM-hY5 wild type plasmid was used and the random mutations were introduced by the short fragment PCR.

B. Overlap extension PCR for generating the mutant pool 4, 5 and 6 PCR products.

Mutant pools 4, 5 and 6 were generated by overlap extension PCR using the wild type pGEM-hY5 plasmid as a template. The short fragment containing the random nucleotides and the long fragment were annealed for 5 cycles, then primers were added and amplified for another 25 cycles.

After re-examining the first cloning strategy (**Figure 4.2**) I designed a second cloning approach and I used a truncated hY5 wild type sequence as a template in order to avoid full length wild type hY5 sequence carryover.

Thus, I generated two different template plasmids for the long and the short fragment in a way that the hY5 full length sequence is disrupted. For the templates I ensured that I picked a single colony for further cloning and overlap extension PCR. The full length hY5 sequence containing the five nucleotide mutated regions for each of the mutant pools was generated by a primer that spanned the truncated part of the hY5 wild type sequence and included an overhang for further overlap extension PCR of the long and the short fragment.

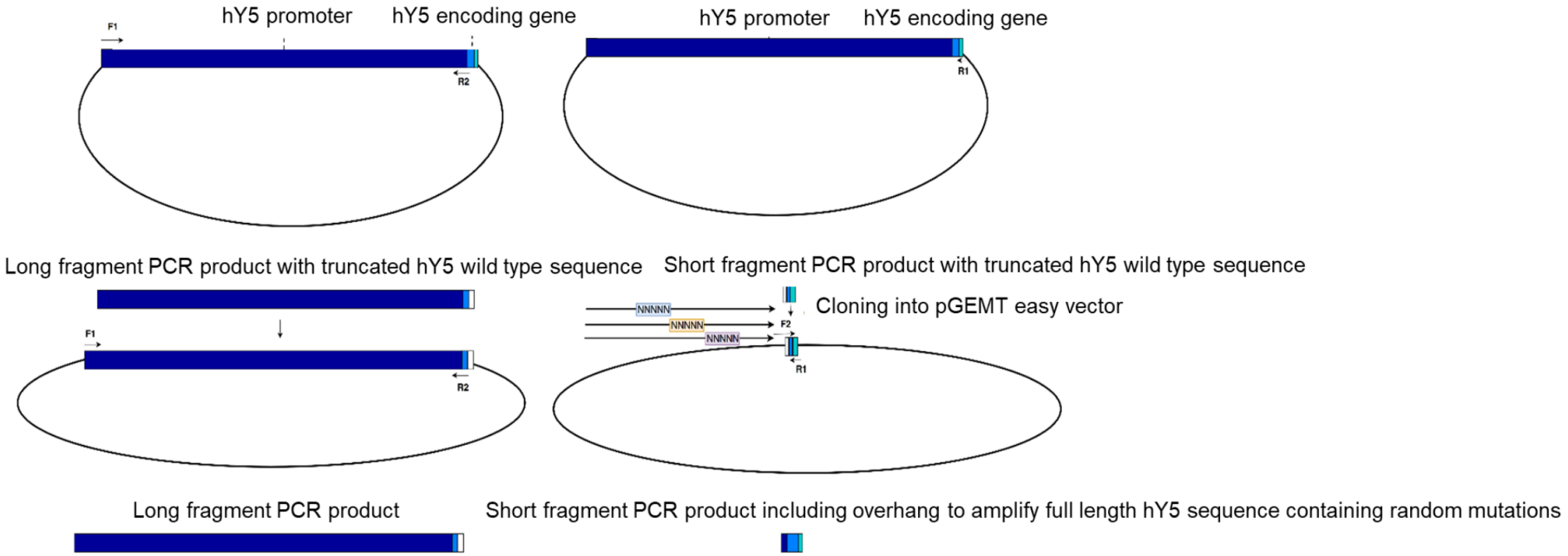
I generated the PCR products of each mutant pool and confirmed them by Sanger sequencing. The PCR products were A-tailed and were cloned into pGEMT easy vector and transformed into bacteria.

The chromatograms of the PCR products showed no obvious contamination and in addition, I sent ten purified plasmid DNA clones of each mutant pools for Sanger sequencing before the generation of the large mutant plasmid pools.

The Sanger sequencing analysis confirmed that the 5' end hY5 mutant pools were generated correctly without any mutations. As previously described I harvested again around 10,000 colonies per library replicate, purified and isolated the large scale 5' end hY5 mutant plasmid pools.

I prepared the plasmid libraries as previously described and sent them for next generation sequencing in order to monitor whether the hY5 wild type sequence bias could be improved.

A


hY5 promoter
hY5 encoding gene
F1
R2
hY5 promoter
hY5 encoding gene
R1
Long fragment PCR product with truncated hY5 wild type sequence
Short fragment PCR product with truncated hY5 wild type sequence
F1
R2
NNNNN
NNNNN
NNNNN
F2
R1
Cloning into pGEMT easy vector
Long fragment PCR product
Short fragment PCR product including overhang to amplify full length hY5 sequence containing random mutations

B

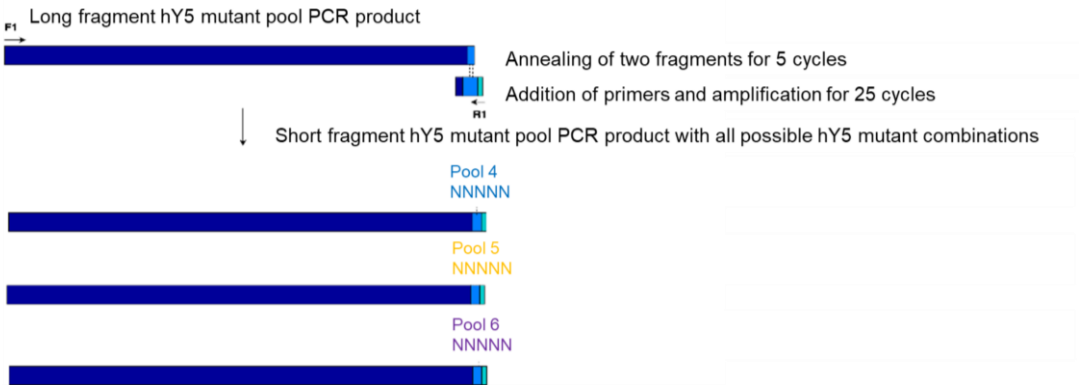


Figure 4.3. Second cloning strategy for 5' end high throughput mutant pool generation.

A. Generation of a long and short PCR fragment with truncated hY5 RNA wild type sequence.

In order to prevent carryover of the full length hY5 RNA wild type sequence in the final overlap extension PCR two separate plasmids containing the long fragment and short fragment each with the truncated hY5 RNA wild type sequence were cloned into pGEMT-easy.

B. Final overlap extension PCR to generate mutant pool 4, 5 and 6 PCR products.

The cloning strategy for the 5' end high throughput mutant pool generation was re-designed and a truncated version of the hY5 wild type sequence was used as a template instead of the full length hY5 RNA wild type sequences. The PCR products of the short and long fragment were cloned into an expression vector and a single colony was picked and checked by Sanger sequencing in order to ensure that the full length hY5 RNA sequence was truncated. The construct containing a truncated version of the hY5 wild type sequence was used to generate the PCR products for the long fragment and short fragment including the random nucleotides for mutant pool 4, 5 and 6. For the final overlap extension PCR the two fragments were annealed for 5 cycles and then the full length forward and reverse primers were added and amplified for 25 cycles to generate the mutant pool 4 (blue), mutant pool 5 (orange) and mutant pool 6 (purple) PCR product.

4.2.3 Bioinformatics analysis of the 5' end hY5 mutagenesis libraries

I constructed all cDNA libraries for the 5' end hY5 high throughput mutagenesis in three replicates and sent the libraries for NGS. The libraries were sequenced on the Illumina HiSeq 2500 at the Earlham Institute (Norwich Research Park, UK). The bioinformatics analysis was carried out by Dr. Simon Moxon and Dr. Archana Singh (John Innes Centre, UK).

Similar to the 3' end hY5 mutagenesis libraries, I prepared 5' end hY5 plasmid libraries and 5' hY5 full length cDNA libraries in order to check if all the mutant hY5 sequences were cloned, transfected and transcribed successfully.

With the analysis of the 5' hY5 plasmid libraries and 5' end hY5 full length cDNA libraries it was possible to normalize the obtained YsRNA reads to some extent. This would provide additional information which 5' end hY5 mutants are cleaved preferentially or not cleaved at all. Combining all the bioinformatics analyses from the 5' end mutagenesis libraries together I could get further insight into the sequence and structural requirements for human Y5 RNA cleavage from the 5' end.

4.2.3.1 Plasmid libraries

I generated the plasmid libraries for the 5' end hY5 mutant plasmid pools as previously described in the material and methods part. I used 5 ng of each mutant pool plasmid DNA as template for PCR. In order to introduce the Illumina adapter sequences and indices for Illumina sequencing I carried out two separate PCR reactions. The Illumina adapter sequences were introduced by PCR using barcoded index primer sequences. The PCR products were separated on an 8% native PAGE gel, the DNA bands excised and purified by ethanol precipitation. I diluted the purified PCR products to 1 ng and used this as a template in order to introduce the Illumina index primer sequences.

The final PCR products were separated on a 8% native PAGE gel, excised and purified by ethanol precipitation (**Figure 4.4, A**).

I measured the concentration of the final plasmid libraries by nanodrop and for quantification 1 µl of each library replicate was loaded on a 8% PAGE gel.

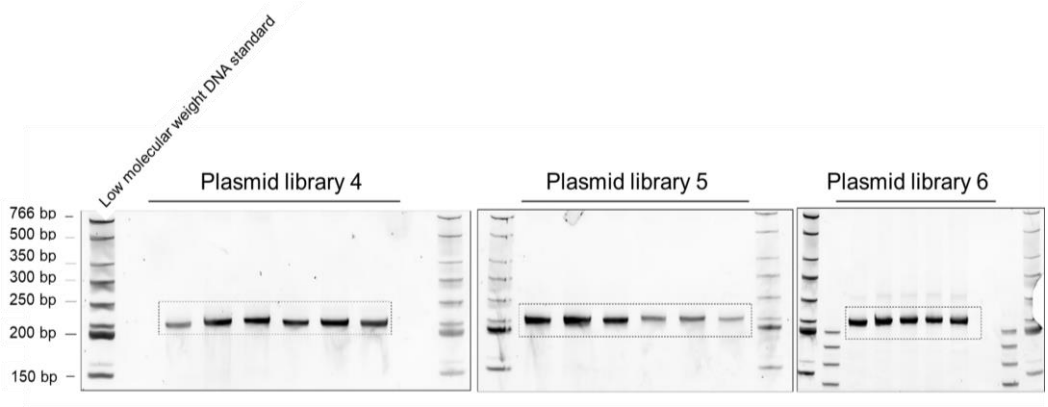
Before the samples were sent for next generation sequencing I quantified all the plasmid library replicates using the Typhoon Image Quant software (**Figure 4.4, B**) The bioinformatics analysis of the 5' hY5 plasmid libraries was performed in a similar way to the 3' end hY5 plasmid libraries. The sequencing reads were mapped to all possible 1024 mutant full length hY5 RNA sequences using PatMaN without allowing any mismatch and further analysed using a custom Perl script. The obtained sequencing reads are summarized in **Figure 4.4, C**.

The bioinformatics analysis of the plasmid libraries showed that for each library replicate we obtained between 300000 and 500000 sequencing reads so for each individual hY5 sequence motifs we expected to get 300 and 500 reads when all the sequence motifs would be equally distributed. When all the obtained reads were mapped to all possible 1024 (4^5) full length hY5 RNA mutant sequences we found all 1024 hY5 RNA mutant sequences represented and the abundance of each sequence motif was more or less equally distributed. From the bioinformatics analysis we could see that there was no particular hY5 RNA mutant sequence overrepresented and that all the individual hY5 mutant sequence motifs were equally present (**Figure 4.4, C**).

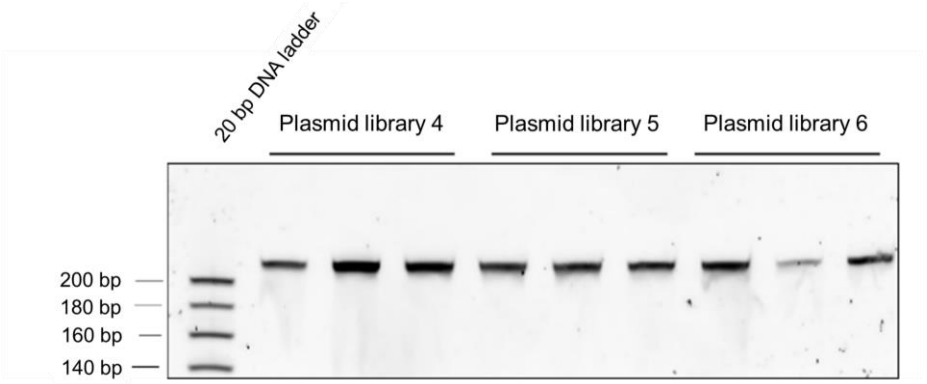
Also, none of the hY5 mutant sequence showed a significant overrepresentation in the plasmid bias libraries and this indicates that the random mutagenesis performed on the 5' end of hY5 RNA was successful.

In conclusion, in the plasmid libraries generated from the mutant pools 4, 5 and 6 all 1024 possible hY5 RNA mutants were present and successfully cloned.

A



B



C

	Total reads	Unique hY5 RNA sequence motifs
PB 4-1	472944	1024
PB 4-2	509178	1024
PB 4-3	473601	1024
PB 5-1	334277	1024
PB 5-2	324255	1024
PB 5-3	347837	1024
PB 6-1	358058	1024
PB 6-2	362396	1024
PB 6-3	423194	1024

Figure 4.4. Plasmid libraries of hY5 libraries 4, 5 and 6.

A. Generation of hY5 plasmid libraries 4, 5 and 6.

For the generation of the plasmid libraries the PCR products were excised and separated on a 8% PAGE gel. The purified PCR products were used as a template to introduce the standard Illumina adapter sequences. The final plasmid DNA libraries were separated on a 8% PAGE gel electrophoresis and bands at 208 nt were excised, eluted and ethanol precipitated.

B. Quantification of hY5 plasmid libraries 4, 5 and 6.

For final quantification 1 µl of each library replicate was loaded on a 8% PAGE gel, stained with SYBR Gold and quantified using the Image Quant software in order to ensure that equal amounts of each plasmid library is sent for NGS.

C. Sequencing information of the 5' end hY5 plasmid libraries.

In the first column all the library replicates of plasmid mutant pool 4, 5 and 6 are listed and for each replicate the abbreviation PB which stands for plasmid bias library was used. The second column shows all the obtained sequencing reads and in the third column all the unique sequence motif for each replicate is listed. In the plasmid libraries generated from plasmid mutant pools 4, 5 and 6 all 1024 possible hY5 RNA mutants were present and successfully cloned.

4.2.3.2 Full length cDNA libraries

I prepared the full length cDNA libraries as previously described. Therefore, 5 µg of total RNA was treated with DNase and purified RNA was reverse transcribed. The Illumina adapter sequences and index primer sequences were added to the hY5 cDNA libraries by PCR (**Figure 4.5, A**).

The PCR products were separated by 8% PAGE gel electrophoresis and quantified using the Image Quant software (**Figure 4.5, B**).

The cDNA full length libraries were sent to the Earlham Institute for next generation sequencing (HiSeq 2500).

The analysis of the 5' end hY5 full length for each library replicate showed that between 100000 and 460000 sequence reads were obtained. After bioinformatics analysis I could show that all the 1024 hY5 mutant combinations were present. This suggested that all pool 4, 5 and 6 hY5 RNA mutants were transcribed successfully and therefore stable.

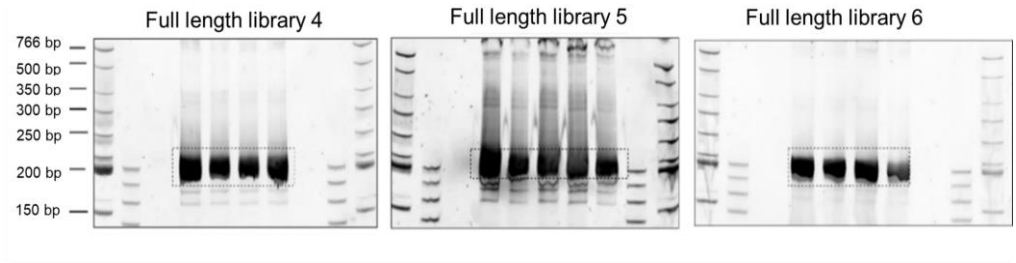
However, I observed that the wild type hY5 sequence seemed to be overrepresented around 3-4 times compared to abundant hY5 RNA mutant transcripts. The obtained sequencing reads for each hY5 library replicates are summarized in (**Figure 4.5, C**).

The analysis of the 5' end hY5 plasmid bias and full length libraries demonstrated that the random mutagenesis on the 5' end of the hY5 RNA was successful and all the hY5 RNA mutants were present. All the hY5 RNA mutant sequences were transcribed successfully and that the generated transcripts were stable.

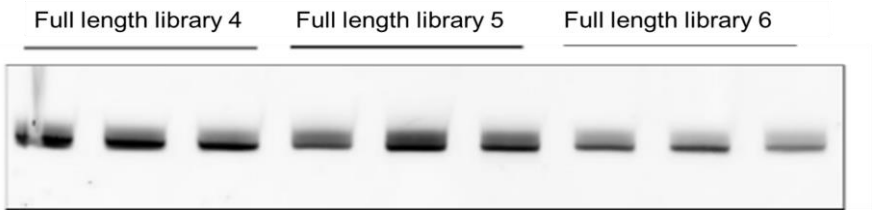
As already been seen in the bioinformatics analysis of the plasmid bias libraries there was no strong selection for any particular hY5 mutant sequence motif and all mutants were represented in the generated mutant pools 4, 5 and 6.

Interestingly, the analysis of the full length libraries showed that the hY5 wild type sequence had around three to four times more transcript level compared to some other abundant hY5 mutant full length transcripts. The reason for this could be that the hY5 wild type sequence was transfected and therefore be transcribed more efficiently or that the wild type full length transcript was more stable compared to the other hY5 mutant full length transcripts.

A



B



C

	Total reads	Unique hY5 RNA sequence motifs
FL 4-1	465275	1024
FL 4-2	281740	1024
FL 4-3	269389	1024
FL 5-1	203573	1024
FL 5-2	269579	1024
FL 5-3	164212	1024
FL 6-1	282315	1024
FL 6-2	303739	1024
FL 6-3	97798	1024

Figure 4.5. Full length cDNA libraries of hY5 library mutant pool 4, 5 and 6.

A. Generation of full length cDNA libraries of hY5 library mutant pool 4, 5 and 6.

The full length cDNA libraries were separated by 8% PAGE gel electrophoresis and PCR products at the size of 200 nt were excised, eluted and precipitated.

B, Quantification of hY5 full length cDNA libraries 4, 5 and 6.

For final quantification 1 µl of each full length cDNA library replicate was loaded on a 8% PAGE gel, stained with SYBR Gold and quantified using the Image Quant software in order to ensure that equal amounts of each plasmid library is sent for NGS.

C. Sequencing information of the 5' end hY5 full length cDNA libraries.

In the first column all the full length cDNA library replicates of mutant pool 4, 5 and 6 are listed and for each replicate the abbreviation FL for full length was used. The second column shows all the obtained sequencing reads and in the third column all the unique sequence motif for each replicate is listed.

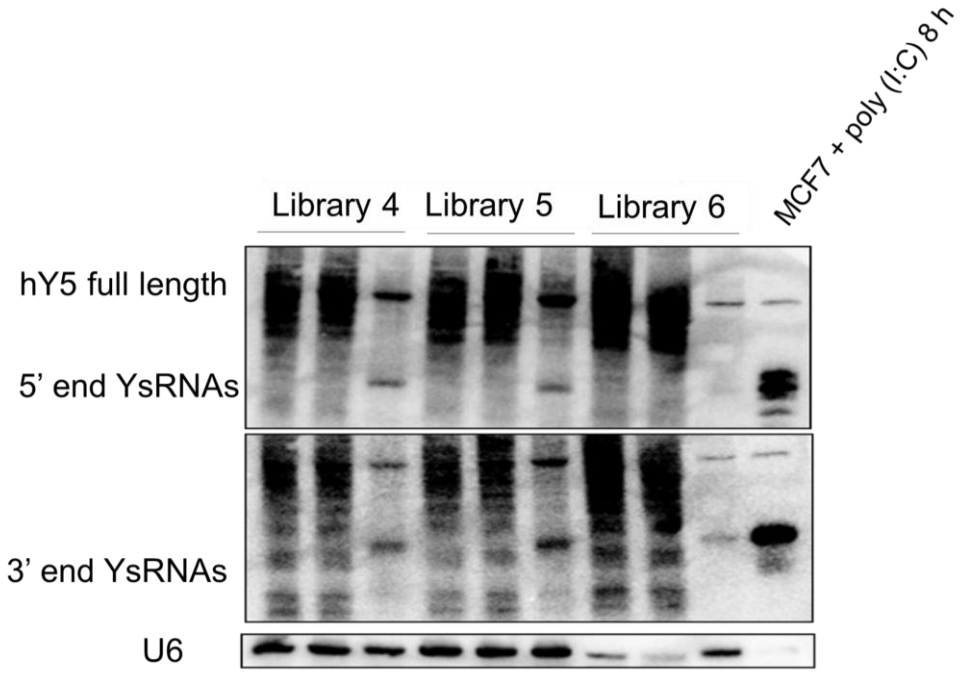
4.2.3.3 Small RNA cDNA libraries

I transfected the mutant plasmid pools into mouse cells and after 24 h the cells were treated with staurosporine for 4 h to induce apoptosis and generate YsRNAs. Total RNA was extracted and transfection and YsRNA production from hY5 mutant pools was checked by Northern blot (**Figure 4.6**, **A**) before cDNA libraries were constructed (**Figure 4.6**, **B and C**) and sent for next generation sequencing.

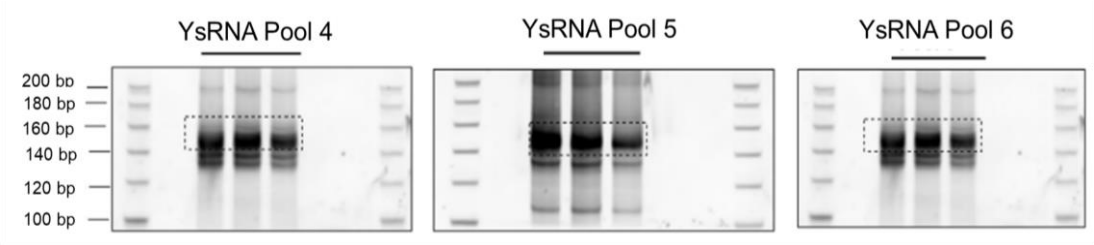
The bioinformatics analysis was carried out by Dr. Simon Moxon. For the YsRNA analysis first the adapter sequence was removed from the 3' end of the sequencing read. Then the reads were mapped to the mouse reference genome. In order to be able to map the obtained reads to the hY5 mutant sequences for each library a reference set containing all possible hY5 mutant sequences was generated. PatMaN was used to map the remaining reads of each library separately to the hY5 mutant sequences while no mismatch was allowed.

From the size class distribution of the YsRNA libraries (**Figure 4.6**, **D**) it can be seen that the size of YsRNAs derived from hY5 library mutant pool 4 and 5 is mainly 29/30 nt whereas hY5 mutant pool 6 Y RNAs generate longer YsRNAs with a length of 30/31 nt. Mutant pool 4 and 5 generate more 28/29 mer YsRNAs whereas mutant pool 6 derived YsRNAs are longer with a length around 30/31 nt.

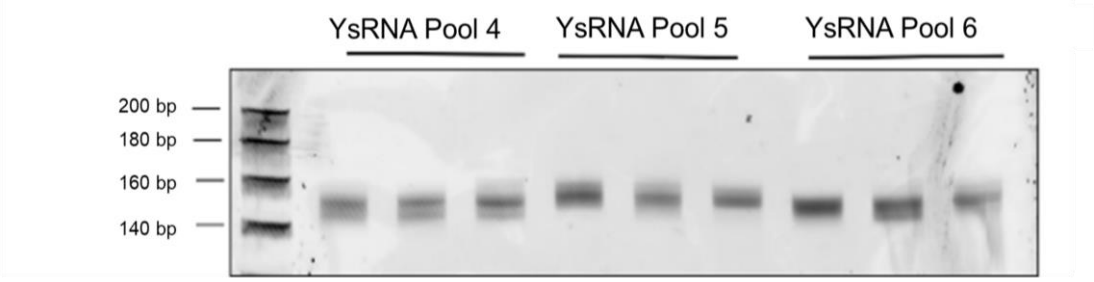
A


MCF7 + poly (I:C) 8 h
Library 4
Library 5
Library 6
hY5 full length
5' end YsRNAs
3' end YsRNAs
U6

B


YsRNA Pool 4
YsRNA Pool 5
YsRNA Pool 6
200 bp
180 bp
160 bp
140 bp
120 bp
100 bp

C


YsRNA Pool 4
YsRNA Pool 5
YsRNA Pool 6
200 bp
180 bp
160 bp
140 bp

D

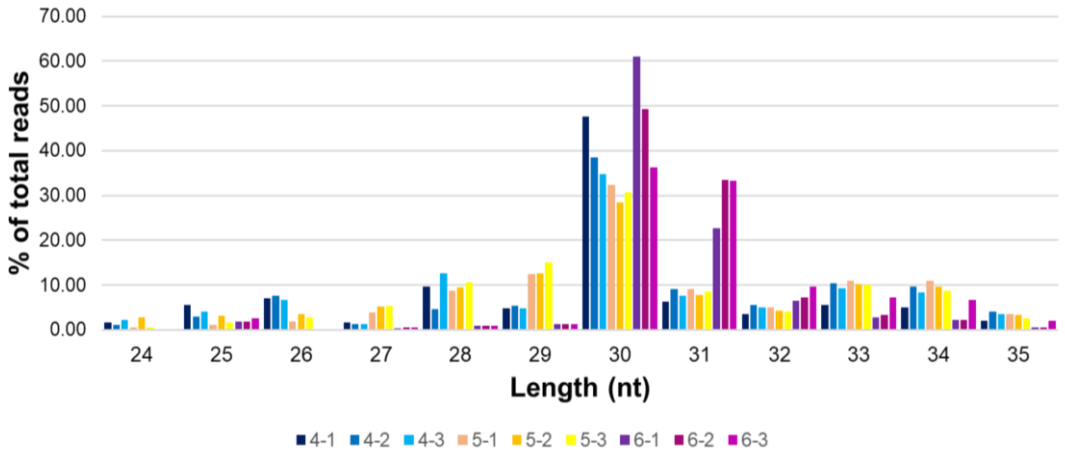


Figure 4.6. Generation of 5' end derived YsRNA cDNA libraries.

A. Northern blot analysis of 5' end derived YsRNAs.

Before cDNA library construction the transfection efficiency and YsRNA generation of mutant pool 4, 5 and 6 was tested by Northern blot analysis. The Northern blot membranes were probed with 3' and 5' end specific hY5 DNA probes and U6 as a loading control. As a size marker total RNA extracted from MCF7 cells treated with poly (I:C) was used.

B. YsRNA library construction.

For YsRNA cDNA library preparation 2 µg of total RNA was used. The YsRNA cDNA libraries were constructed using Illumina adapters. The PCR products were separated on a 8% PAGE gel and PCR products with sizes of 150 nt were excised, eluted and ethanol precipitated. As a size marker a 20 bp DNA ladder was used.

C. Quantification of YsRNA cDNA libraries.

For quantification 1 µl of each library replicate was loaded on a 8% PAGE gel. The gel was stained with SYBR Gold and the bands were quantified using the Typhoon Image Quant software.

D. Size class distribution of YsRNAs from each hY5 library mutant pool replicate.

Mutant pool 4 and 5 YsRNAs (shown in blue and yellow) are mainly 30 nt whereas mutant pool 6 hY5 RNA mutants mainly generate YsRNAs of 30 nt and longer YsRNAs of 31 nt.

4.2.4 Library 4 hY5 RNA mutants generate wild type sized YsRNAs

The bioinformatics analysis was carried out in a similar way to the 3' end hY5 YsRNA analysis. The size class distribution of YsRNA sequencing reads demonstrated that hY5 library 4 mutant YsRNA reads were mainly 30 nt as seen for the wild type hY5 RNA. Interestingly, it was noted that hY5 RNA was cleaved at different positions specifically at the 5' end and generated YsRNA fragments with various sizes.

The YsRNAs derived from each hY5 mutant pool 4 library replicates were ranked by abundance. In order to validate the results from the next generation sequencing I performed some RNA secondary structure analysis of 30 library 4 hY5 RNA mutant sequences with the most abundant and least abundant YsRNAs using RNAfold. In addition, I carried out sequence logo analysis of the thirty most abundant library 4 hY5 mutant sequences using WebLogo3.

From the secondary structure analysis I found that the overall structure of the most abundant hY5 library 4 mutants is more or less similar to the wild type hY5 RNA (**Figure 4.7, C**).

The sequence logo analysis clearly showed that there was quite a strong selection for the wild type hY5 RNA sequence motif UGGGU (**Figure 4.7, B**).

This indicates that the region mutated in hY5 library 4 seems to be important for Y RNA cleavage. Intriguingly, the mutated region of library 4 is part of the 8 nt conserved sequence motif (5' GUUGUGGG 3') that was shown to be required for DNA replication and was recently reported to be involved in triggering apoptosis in 5' end derived hY5 fragments (Chakrabortty et al., 2015).

The cloned some library 4 hY5 RNA mutants that generated the most YsRNAs (

Figure 4.7, **A**) individually into an expression vector and tested their expression level by Northern blot analysis.The Northern blot results of the library 4 hY5 RNA mutants that resulted in the five most abundant YsRNA reads (Figure 4.7,

D)confirmed that hY5 RNA mutant sequences produced different sized 5' end derived hY5 RNA fragments and hY5 RNA cleavage mainly happened between position A30 and A31 (

Figure 4.7, red arrows). In the hY5 L4-6 mutant the hY5 RNA sequence UGGGU was mutated to CAACC. Intriguingly this mutant folded completely differently compared to the wild type and did not generate YsRNAs either from the 3' end or the 5' end of hY5 RNA (Figure 4.7, C).

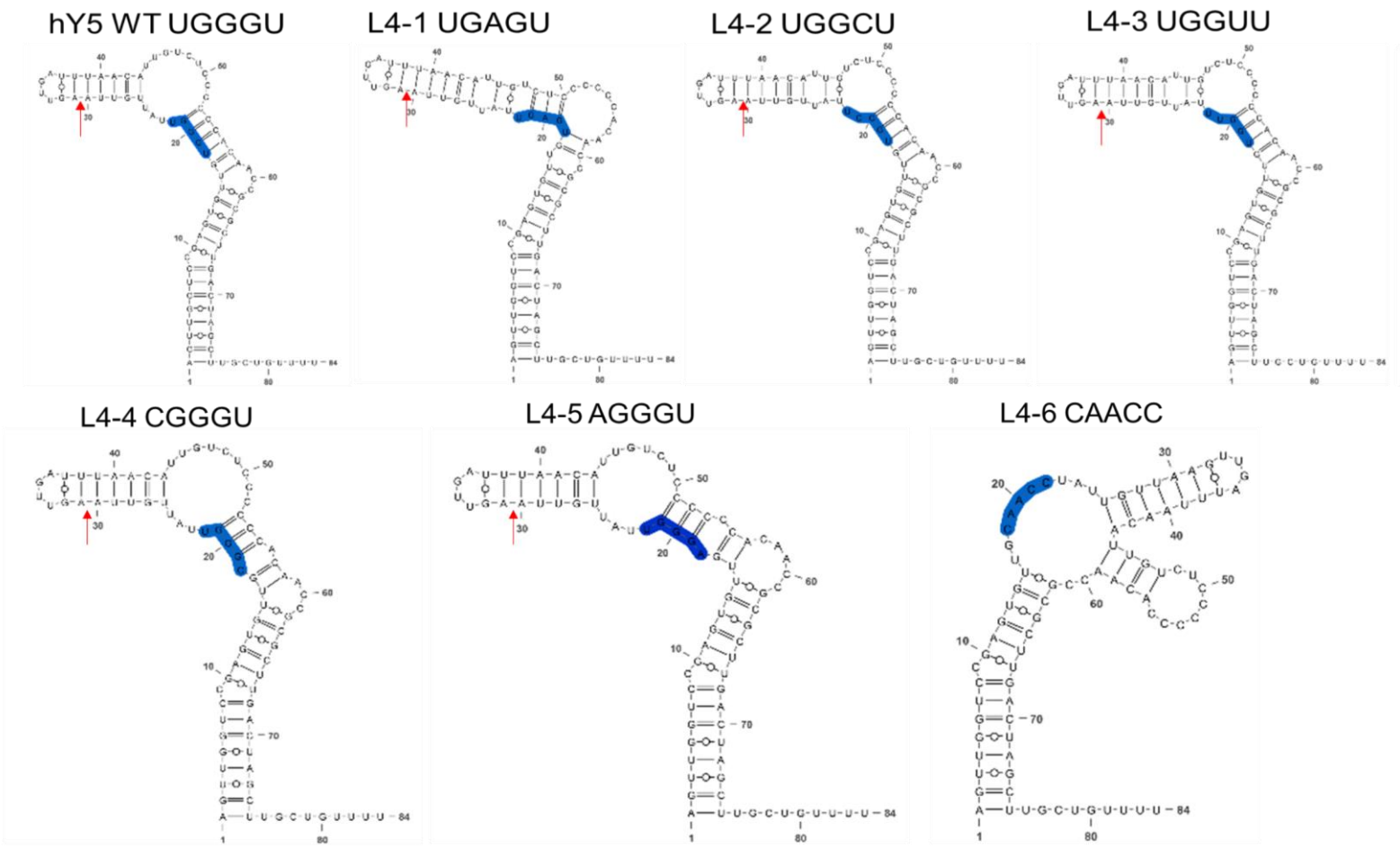
A

Mutant	Abundance
L4-1 (UGAGU)	20128
L4-2 (UGGCU)	18032
L4-3 (UGGUU)	14158
L4-4 (UGGGU)	12424
L4-5 (CGGGU)	10659
L4-6 (CAACC)	12

B


5′
1
2
3
4
5
3′
weblogo.berkeley.edu

C


hY5 WT UGGGU
L4-1 UGAGU
L4-2 UGGCU
L4-3 UGGUU
L4-4 CGGGU
L4-5 AGGGU
L4-6 CAACC

D

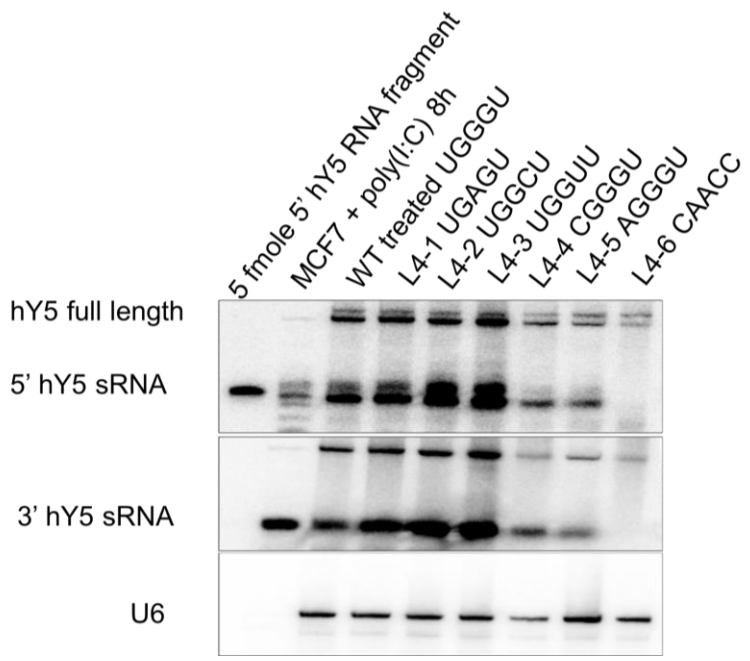


Figure 4.7. Library 4 hY5 RNA mutants generate wild type-sized YsRNAs.

A. Abundances of validated mutant pool 4 hY5 sRNA mutants.

The abundances of each hY5 library 4 mutant with the most abundant YsRNAs were normalized across all the three replicates.

B. Sequence logo analysis of 30 most abundant hY5 library 4 mutants that generated the most YsRNAs using WebLogo 3.

The sequence logo UGGGU seems to be strongly preferred compared to other L4 hY5 mutant sequences.

C. Predicted secondary structures of the L4 hY5 RNA mutants with the most abundant YsRNA reads and one low abundant L4 hY5 mutant.

The secondary structures were predicted using RNAfold and were visualized using Varna. The L4 hY5 RNA mutants the five most abundant and one low abundant YsRNA reads. The L4 hY5 RNA sequences fold similar to the wild type hY5 RNA structure. The red arrows indicate the observed cleavage positions.

D. Northern blot analysis of the library 4 hY5 RNA mutants with the five most abundant library 4 hY5 RNA mutants and one low abundant YsRNA reads.

Total RNA extracted from human MCF7 cells treated with poly (I:C) and a synthetic 3' end derived hY5 fragment of 31 nt were used as a size marker for full length hY5 RNA and wild type-sized YsRNAs.

Total RNA was extracted from the five most abundant L4 hY5 mutants and one low abundant library 4 hY5 mutant and analysed by Northern blotting. The Northern blot was probed with the 3' and 5' YsRNA probes. The blot was re-probed with U6 as a loading control. The YsRNAs produced from the five most abundant hY5 library 4 mutants were at the same size as wild type hY5 YsRNAs.

4.2.5 Library 5 hY5 RNA mutants produce wild type sized YsRNAs

The bioinformatics analysis of YsRNA sequencing reads showed that hY5 library mutant pool 5 derived YsRNAs are mainly 30 nt reads as seen for hY5 library 4 mutants and the wild type hY5 RNA. The library 5 mutant pool derived YsRNAs were ranked by abundance and I analyzed 30 library 5 hY5 RNA mutants with the most abundant and least abundant YsRNA using RNAfold. In addition, I carried out sequence logo analysis of 30 library 5 hY5 RNA mutant sequences that generated the most YsRNAs using WebLogo3.

The secondary structure analysis showed that the overall structure of the most abundant hY5 library 5 mutants is more or less similar to the wild type hY5 RNA. However, for some of the L5 hY5 RNA mutants the secondary structure changed in way that an additional internal loop is formed further upstream close to the introduced mutations of L5 as could be observed for L5-2 UAUGG and L5-4 UAUGA mutants (**Figure 4.8, C**).

The sequence logo analysis showed that there is a preference for the RNA sequence UAU at the first three positions of library 5 mutated region (**Figure 4.8, B**)

At position 5 of library 4 The sequence logo of library 4 which was shown in Figure 4.7 revealed that uridine at the 5^{th} position of the sequence logo is preferred.

Also, I did a multiple sequence alignment of all human Y RNAs and I could clearly show that the motif UUAU is conserved among hY5, hY4 and hY1 but not hY3, whereas the U at the third position is highly conserved among all human Y RNAs.

I cloned the hY5 mutant pool 5 library mutants that generated the five most abundant YsRNAs (**Figure 4.8**, **A**) individually into pGEMT easy vector and tested their expression level by Northern blot analysis.

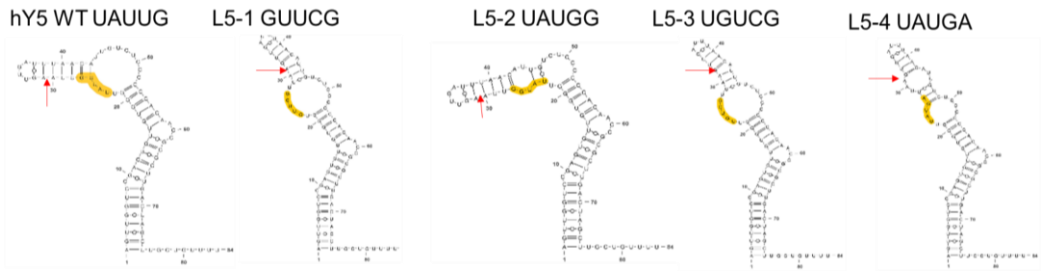
The Northern blot analysis of mutant pool 5 hY5 RNA mutants revealed that all the tested hY5 RNA mutants generated wildtype-sized YsRNAs (**Figure 4.8, C**).

A

Mutant	Abundance
L5-1 (GUUCG)	1791
L5-2 (UAUGG)	1641
L5-3 (UGUCG)	1218
L5-4 (UAUUG)	1140
L5-5 (UAUGA)	988

B

C



D

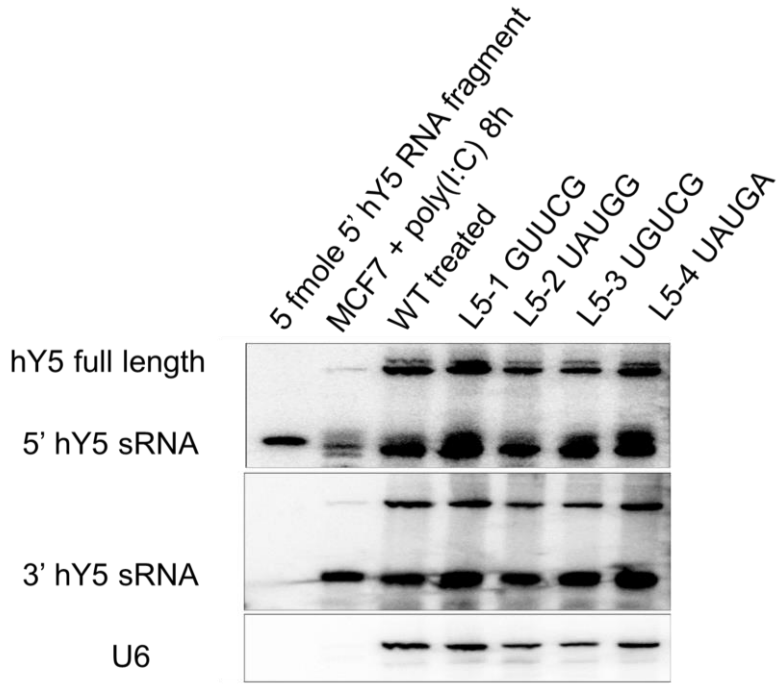


Figure 4.8. hY5 library 5 mutant Y RNAs produce wild type sized YsRNAs.

A. Abundances of validated mutant pool 5 hY5 sRNA mutants.

The abundances of each hY5 library 5 mutant with the most abundant YsRNAs were normalized across all the three replicates.

B. Sequence logo analysis of library 5 hY5 RNA mutants with the 30 most abundant YsRNA reads.

The sequence logo analysis was carried out using Weblogo3. The sequence logo analysis shows sequence preference for the motif UAUCG for 30 mer YsRNAs.

C. Predicted secondary structures of L5 hY5 RNA mutants with the four most abundant YsRNAs.

The secondary structures were predicted using RNAfold and visualized using Varna.

The L5 hY5 RNA sequences that resulted in the four most abundant YsRNAs fold more or less similar to the wild type hY5 RNA structure, but some L5 hY5 RNA mutants have an additional internal loop.

D. Northern blot analysis of L5 hY5 RNA mutants with the five most abundant library 5 hY5 RNA mutants.

Total RNA extracted from human MCF7 cells treated with poly (I:C) and a synthetic 3' end derived hY5 fragment of 31 nt were used as a size marker for full length hY5 RNA and wild type-sized YsRNAs. Total RNA was extracted from the L5 hY5 RNA mutant sequences that resulted in the five most abundant YsRNAs and compared to the wild type hY5 RNA derived YsRNAs by Northern blotting. The Northern blot was probed with the 3' and 5' YsRNA probes. As a loading control the Northern blot membrane was re-probed with U6. The five most abundant hY5 library mutants that were analysed produced wild type-sized hY5 YsRNAs both from the 3' end and the 5' end of the hY5 RNA.

4.2.6 Library 6 hY5 RNA mutants generate YsRNAs at the same size as wild type

The size class distribution of YsRNAs from library 6 revealed that the length of hY5 mutant pool 6 YsRNAs is mainly 30 nt (**Figure 4.6**, **D**). Interestingly, hY5 library 6 mutants also generated longer YsRNA fragments of 31 nt that could not be detected in hY5 mutant pool 4 and 5.

Due to the fact that hY5 library 6 mutants mainly produced YsRNAs with 30/31 nt and not YsRNAs with a length of 32 nt as we expected, the sequence information of the fourth and fifth nucleotide for the 30 mer YsRNAs and the fifth nucleotide for the 31 mer YsRNAs of the mutated five nucleotide hY5 library 6 motif got lost. Even with the bioinformatics analysis of the hY5 library 6 full length libraries we could not be sure from which mutant motif the YsRNAs of 30/31 nt were derived.

The YsRNAs from library 6 were ranked by abundance (**Figure 4.9**, **A**) and I folded 30 library 6 hY5 RNA mutant sequences with the most abundant YsRNA reads using RNAfold. I also performed sequence logo analysis of all these L6 hY5 RNA mutant sequences using WebLogo3.

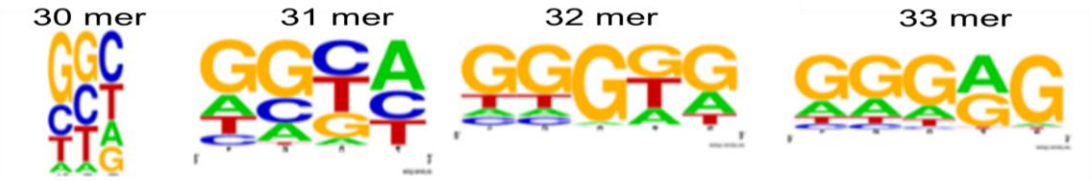
The predicted secondary structures of hY5 library 6 mutants folded in a different way than the wild type hY5 RNA (**Figure 4.9, C**). The sequence logo analysis showed that for 32/33mer library 6 mutant YsRNAs there seems to be a preference for guanosine at each position of the mutated region (**Figure 4.9, B**).

The Northern blot analysis demonstrated that the pool 6 mutants produced YsRNAs with 30 nt and 31 nt from the 5' end (**Figure 4.9**, **D**).

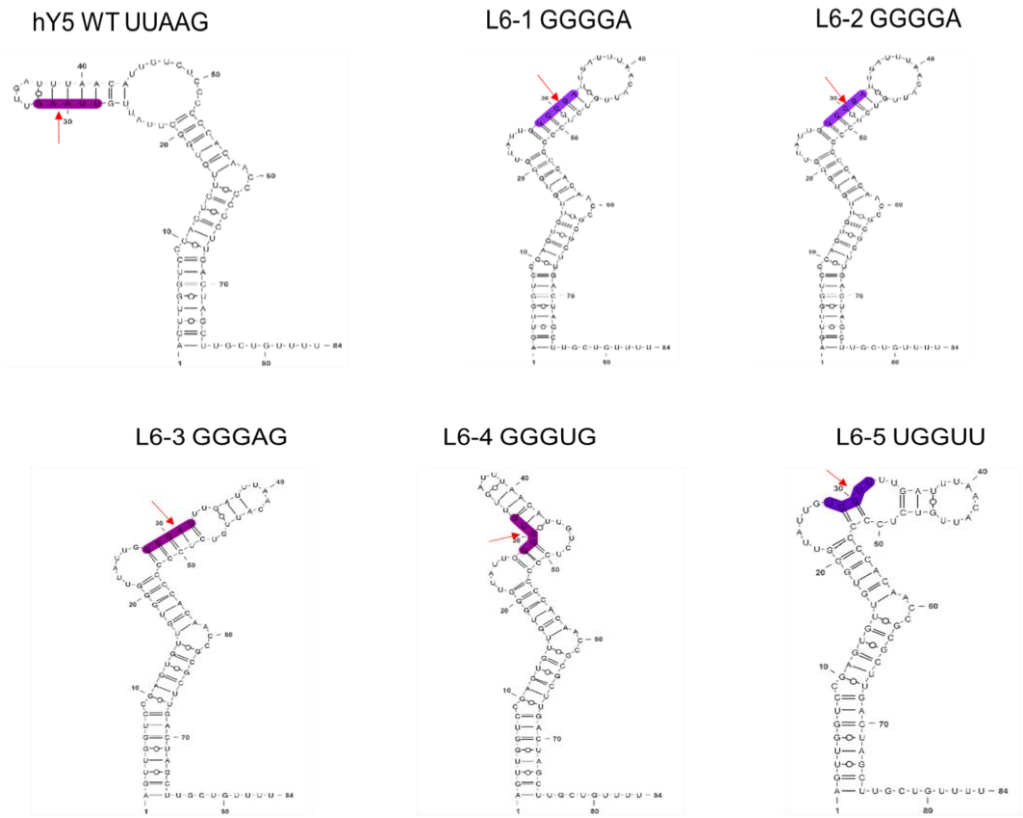
A

Mutant	Abundance
L6-1 (GGGGA)	1181
L6-2 (GAGGG)	932
L6-3 (GGGAG)	913
L6-4 (GGGUG)	863
L6-5 (UGGUU)	645

B



C



D

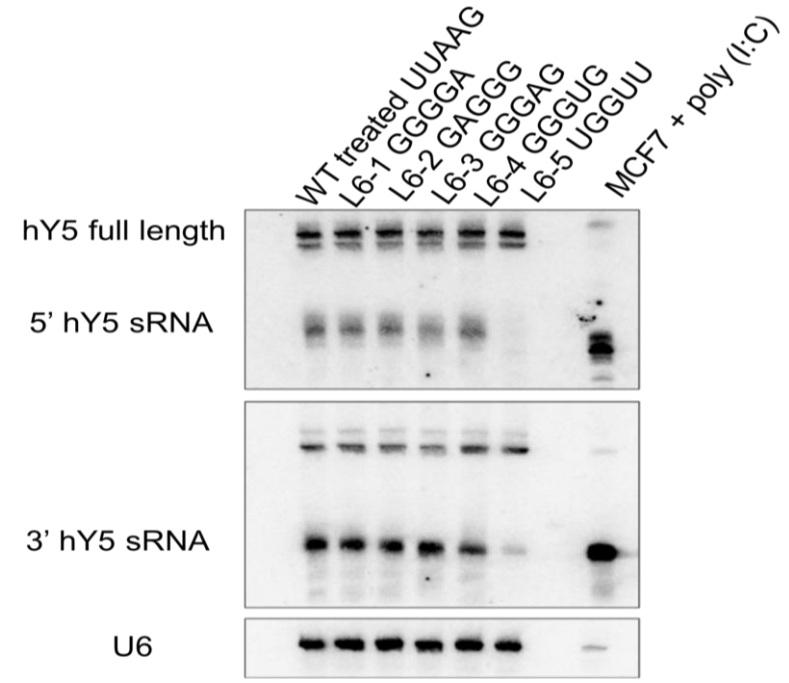


Figure 4.9. Library 6 mutant Y RNAs produce wild type sized YsRNAs.

A. Abundances of validated mutant pool 6 hY5 RNA mutants.

The abundances of L6 hY5 RNA mutants that resulted in the most abundant YsRNAs were normalized for each library replicate and the abundance of L6 hY5 RNA mutants with the four most abundant 32 mer YsRNAs are listed.

B. Sequence logo analysis of 30 L6 hY5 RNA mutant sequence that generated the most YsRNAs.

The sequence logo analysis was performed using WebLogo 3. For L6 hY5 RNA mutants no sequence motif preference could be observed.

C. Predicted secondary structures of L6 hY5 RNA mutants with most abundant YsRNAs.

The secondary structure analysis was performed using RNAfold and the structures were visualized by Varna. The L6 hY5 RNA mutants with the most abundant fold in a different way than the wild type hY5 RNA.

D. Northern blot analysis of the five most abundant library 6 hY5 RNA mutants.

Total RNA extracted from human MCF7 cells treated with poly (I:C) and a synthetic 3' end derived hY5 fragment of 31 nt were used as size markers for full length hY5 RNA and wild type-sized YsRNAs. Total RNA was extracted from the five most abundant L6 hY5 mutants and compared to the wild type hY5 RNA by Northern blotting. The Northern blot was probed with the 3' and 5' YsRNA probes. As a loading control the Northern blot membrane was probed with U6. The five most abundant hY5 library mutants that were analysed produced wild type-sized hY5 YsRNAs both from the 3' end and the 5' end of the hY5 RNA.

4.2.7 Is the loop L2b important for human Y5 RNA cleavage?

Bioinformatics analysis and sequence logo analysis of the mutant pool 4 and 5 YsRNAs indicated that the region around the loop L2b (**Figure 4.10, A**) might be important for human Y5 RNA cleavage from the 3' and 5' end.

Interestingly, the upper stem domain of Y RNA is characterized by a highly conserved A/GUG-CAC-U nucleotide sequence motif that was previously reported to be required for chromosomal DNA replication initiation (Christov et al., 2006; Gardiner et al., 2009; Wang et al., 2014). Interestingly, in nematode sbRNAs the penta-nucleotide motif UUAUC motif was identified which is part of the internal loop L2b. Thus, this region of hY5 RNA might be essential for the initiation of DNA replication as it was already observed in nematode sbRNAs and Dm1 in *Drosophila melanogaster*. The region which was shown to be essential of DNA replication initiation is shown in light green and is close to the internal loop L2b.

4.2.7.1 The uridine at position 22 essential for Y RNA cleavage from the 5' end

In this part of the work I investigated if the wild type RNA sequence UGGGU is necessary for 3' and 5' end hY5 RNA cleavage. The sequence logo analysis of mutant pool 4 YsRNAs showed a strong preference for U at the last position (**Figure 4.7, B**).

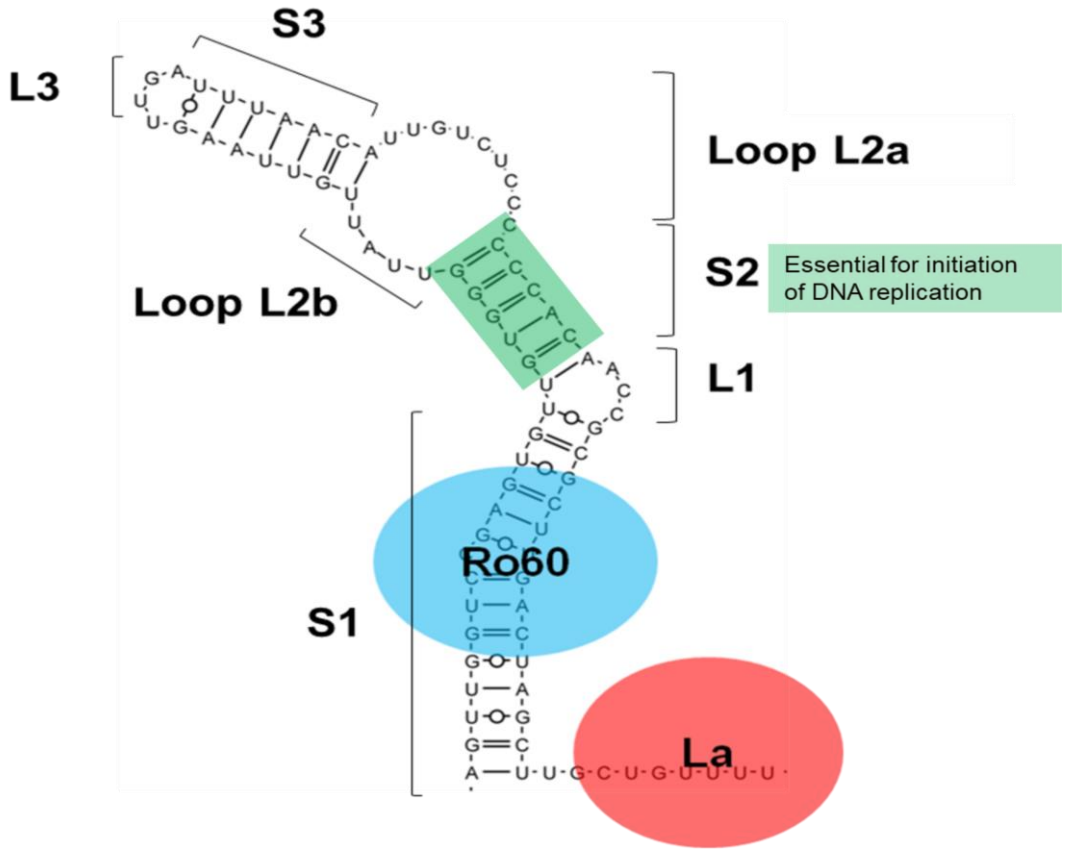
Interestingly, in the multiple sequence alignment of all human Y RNAs I observed that uridine at the first position, guanine at the second position and uridine at the fifth position are conserved in all the human Y RNAs (**Figure 4.10, B**)**.**

In order to check if the uridine at the fifth position of the RNA sequence logo is essential, I replaced the U by A, C and G. I generated the hY5 RNA mutants UGGGC and UGGGA and UGGGG and looked at the YsRNA accumulation by Northern blot analysis. Interestingly, the mutant UGGGC folded exactly in the same manner as the wild type sequence (**Figure 4.10, C**). However, when the YsRNAs that were derived from the hY5 RNA mutant UGGGC were compared with the wild type hY5 derived YsRNAs by Northern blotting I could show that Y RNA cleavage was clearly affected and impaired at the 5' end of the hY5 RNA.

From the 3' end of hY5 RNAs Y RNA cleavage still occurred in the mutants but Y RNA cleavage was decreased in the substitution mutants compared to the wild type hY5 RNA.

From these results it can be concluded that the uridine at position 22 of the conserved Y RNA sequence motif UGGGU is essential for 5' end hY5 RNA cleavage and might be involved in 3' end hY5 RNA cleavage.

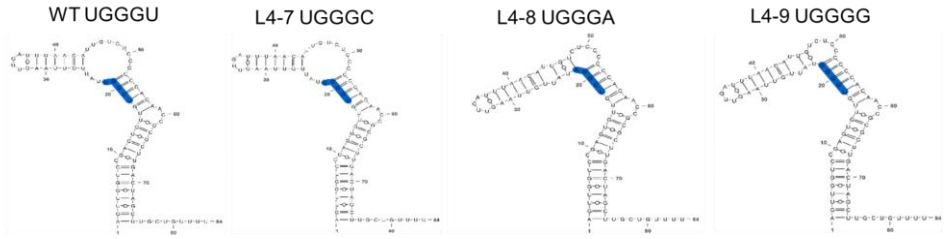
A



B

```
hY1             GGCTGGTCCGAAGGTAGTGAGTTATCTCAATTGATTGTTCACAGTCAGTTACAGATCGAA
hY3             GGCTGGTCCGAGTGCAGTG-GTGTTTACAACTAATTGATCACAACCAGTTACAGATT---
hY4             GGCTGGTCCGATGGTAGTGGGTTATCAGAACTTATTAACATTAGT--GTCACTAAAG---
hY5             AGTTGGTCCGAGTGTTGTGGGTTATTGTTA--AGTTGATT--------TAACA-------
                 * ********  *  *** **  *    *    **          * **
```

C



D

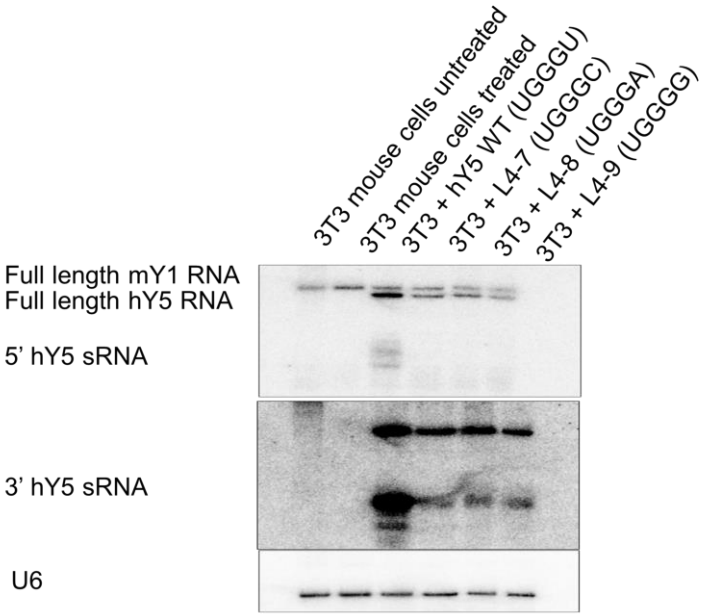


Figure 4.10. The uridine at position 22 is essential for 5' end hY5 RNA cleavage and might be involved in 3' end hY5 RNA cleavage.

A. Predicted secondary structure of human Y5 RNA including characteristic secondary structure features.

The human Y5 RNA has a characteristic stem loop structure. The Ro60 protein binds to a C-bulge within the highly conserved stem S1, whereas the La protein interacts with the uridine-stretch at the 3' end. The secondary structure of hY5 contains stem regions (S1, S2 and S3) as well as several internal loop regions or smaller bulges (L1, L2a, L2b, L3).

B. Multiple sequence alignment of the 5' end of all human Y RNAs using the ClustalW algorithm.

The multiple sequence alignment shows that the pool 4 mutated region TGGGT is conserved at the first, second and last position.

C. Secondary structures of hY5 RNA with the wild type sequence motif UGGGU and substitutions hY5 RNA mutants.

In hY5 RNA mutant pool 4 sequences the uridine residue at position 22 was substituted by adenine, cytosine and guanosine. The secondary structures were folded using RNAfold and visualized by Varna.

D. Northern blot analysis of mutant pool 4 YsRNAs derived from hY5 substitution mutants UGGGA, UGGGC and UGGGG compared to the wild type YsRNA accumulation.

Mouse 3T3 cells were transfected with plasmid DNA constructs encoding mutant pool 4 hY5 RNA sequences with the sequence motifs UGGGA, UGGGC and UGGGG and the wild type hY5 RNA sequence motif UGGGU. Total RNA was extracted and the expression of mutant pool 4 derived YsRNAs compared to wild type hY5 RNA derived YsRNAs by Northern blotting. As a size marker a synthetic 5' end derived hY5 fragment and RNA extracted from MCF7 cells treated with poly (I:C) were loaded. The Northern blot membrane was hybridized using the 5' and 3' end hY5 RNA probes. U6 was used as a loading control.

4.2.7.2 Role of UUAU motif at position 22-25 for 3' and 5' end hY5 RNA cleavage

Intriguingly, further secondary structure and northern blot analysis of library 1 hY5 RNA mutants (described in the previous chapter 3.2.2) also revealed that the loop L2b at the position 22-26 of hY5 RNA (shown in turquoise in **Figure 4.11**, **A**) might contribute to YsRNA generation from the 3' and 5' end.

For this analysis, I compared the abundance of YsRNAs derived from mutant pool 1 to the abundance of mutant pool 1 hY5 full length sequences. The YsRNAs with the highest number of abundance were ranked as number one (**Figure 4.11**, **B**).

First, I predicted the secondary structures of the 30 most abundant L1 hY5 mutants that produced a significant number of YsRNAs compared to full length transcripts using RNAfold (**Figure 4.11**, **C and D**).

I could show that the secondary structures folded in a similar way as the wild type hY5 RNA (**Figure 4.11**, **C and D**). I performed a detailed secondary structure analysis and is summarized in **Figure 4.11**, **E.**

From the secondary structure analysis (**Figure 4.11, E**) it can be seen that the L1 hY5 mutants that generated a small amount of YsRNAs compared to the full-length transcript showed a different secondary structure than the wild type hY5 RNA.

The hY5 RNA library 1 mutants that showed a high accumulation of YsRNAs compared to the full length hY5 RNA all of these mutants had an internal loop L2b consisting of 4 nt with the sequence UUAU whereas hY5 library 1 mutants that produced a low amount of YsRNAs compared to the full length hY5 RNA all had a shortened internal loop L2b of 3 nt. Additionally, it was noted that in mutants that showed a high accumulation of YsRNAs the internal loop L2a where the 3' hY5 RNA cleavage occurs was 9 nt. In the hY5 RNA library 1 mutants with low amount of YsRNAs this internal loop L2a was enlarged by 1 nt to 10 nt in total. Also, the stem regions S2 and S3 seemed to be less stable than in the hY5 RNA mutants that generated a lot of YsRNAs.

Notably, the main difference between the mutants that had a high accumulation of YsRNAs and the mutants that produced a low amount of YsRNAs was that the internal loop L2b in the less efficient L1 hY5 RNA mutants got shortened from four nucleotides to three nucleotides.

The sequence logo analysis that was performed for mutant pool 4 and 5 and the change of secondary structure introduced by L1 hY5 RNA mutations was the rationale for further investigation into the importance of the internal L2b loop.

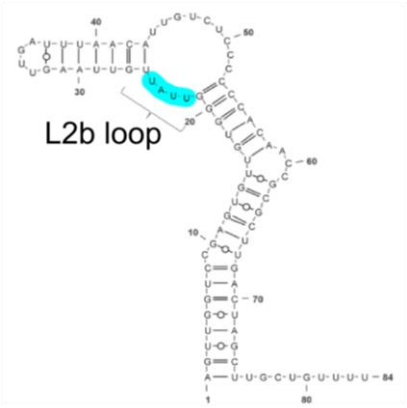
In order to validate the results from the secondary structure analysis I cloned three L1 hY5 mutants that produced a good amount of YsRNAs and three L1 hY5 mutants that produced only a small amount of YsRNAs compared to the full length hY5 mutant transcript (**Figure 4.11**, **A**) individually and I transfected the constructs into mouse cells. The cells were treated to induce Y RNA cleavage. I tested the YsRNA accumulation by Northern blot.

The Northern blot analysis and relative quantification of the Northern blots by densitometry confirmed the bioinformatics analysis and showed the difference in YsRNA accumulation between the hY5 mutant sequences that produced a good amount of YsRNAs and the mutants that produced little or no YsRNAs compared to the full length hY5 RNA (**Figure 4.11**, **F, G, H and I**). When the ratio between the produced mutant YsRNA levels and the full length hY5 RNA mutant transcript levels calculated was compared I could show that the library 1 hY5 RNA mutants with the internal loop L2a intact are more efficiently processed than the library 1 hY5 RNA mutants in which the internal loop L2a got shortened by the introduced mutations. Also, for the 5' end derived YsRNAs I observed a decrease of YsRNA accumulation when the internal loop L2a got shortened (**Figure 4.11, I**).

In conclusion, both the 3' and 5' end hY5 RNA cleavage seemed to be impaired in the hY5 mutants with a shorter internal loop L2b (**Figure 4.11**, **G and I**). These results together with the secondary structure predictions of these mutants suggested that the internal loop L2b might contribute to 3' and 5' hY5 RNA cleavage.

,

A

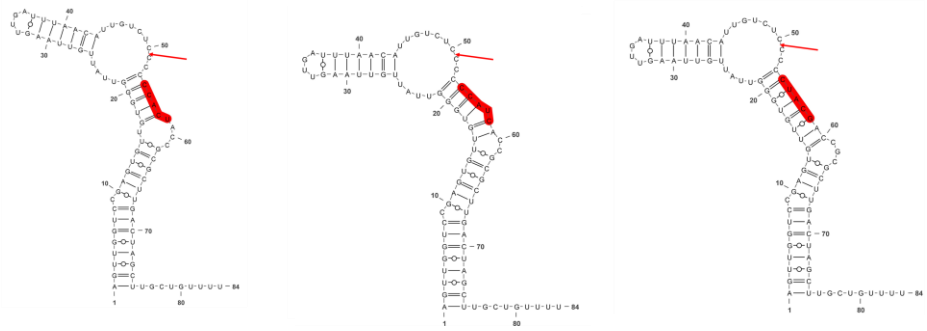


B

Mutant	Abundance full length transcript	Rank full length transcript	Abundance YsRNA	Rank YsRNA
L1-CCACU	2113	215	600	4
L1-CCAUC	1994	223	274	13
L1-CUACG	1788	208	208	15
L1-GAUCC	6333	5	3	52
L1-GAACC	5610	14	2	585
L1-GACCU	5579	19	1	582

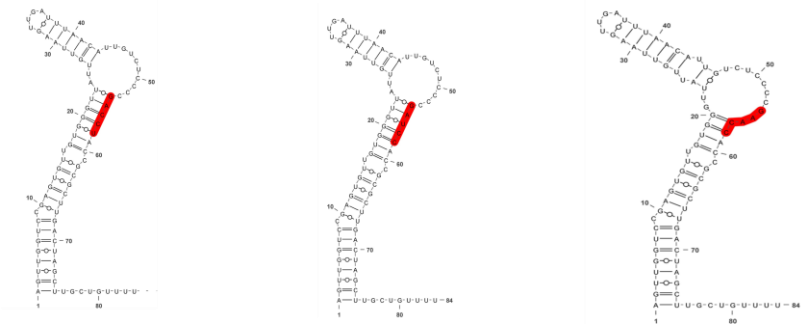
C

L1 CCACU L1 CCAUC L1 CUACG



D

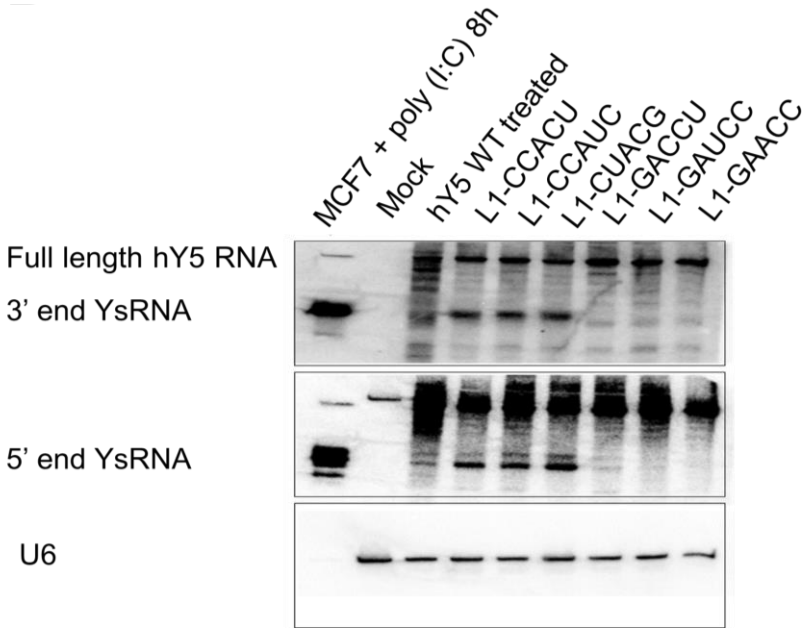
L1 GACCU L1 GAUCC L1 GAACC



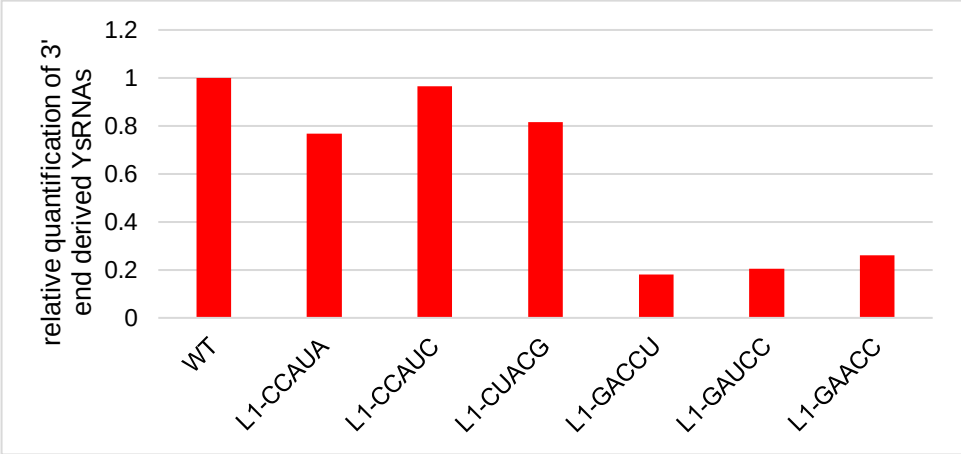
E

RNA motif	Stem S3	Size of internal loop L2a	Size of internal loop L2b	Stem S2
WT CCACA	7 nt 1x G-C 5x A-U 1x G-U	9 nt 4x U 2x G 3x C	4 nt 3x U 1x A	6 nt 4x G-C 2x A-U
L1-CCACU	7 nt 1x G-C 6x A-U	9 nt 4x U 1x G 4x C	4 nt 3x U 1x A	5 nt 4x G-C 1x A-U
L1-CCACU	7 nt 1x G-C 6x A-U	9 nt 4x U 1x G 4x C	4 nt 3x U 1x A	7 nt 4x G-C 2x A-U 1x U-bulge
L1-CUACG	7 nt 1x G-C 6x A-U	9 nt 4x U 1x G 4x C	4 nt 3x U 1x A	7nt 4x G-C 2x G-U 1x A-U
L1-GACCU	6 nt 1x G-C 5x A-U	10 nt 4x U 1x G 5x C	3 nt 2x U 1x A	7 nt 3x G-C 2x U-G 2x A-U
L1-GUUCC	8 nt 1x G-C 6x A-U 1x G-U	10 nt 4x U 1x G 5x C	3 nt 2x U 1x A	7 nt 2x G-U 3x G-C 2x A-U
L1-GAACC	10 nt 4x non - canonical 2x A-U 1x U-bulge 3x G-U	10 nt 6x C 1x U 2x G 1x A	3 nt 2x U 1x A	4 nt 2x G-C 2x non - canonical

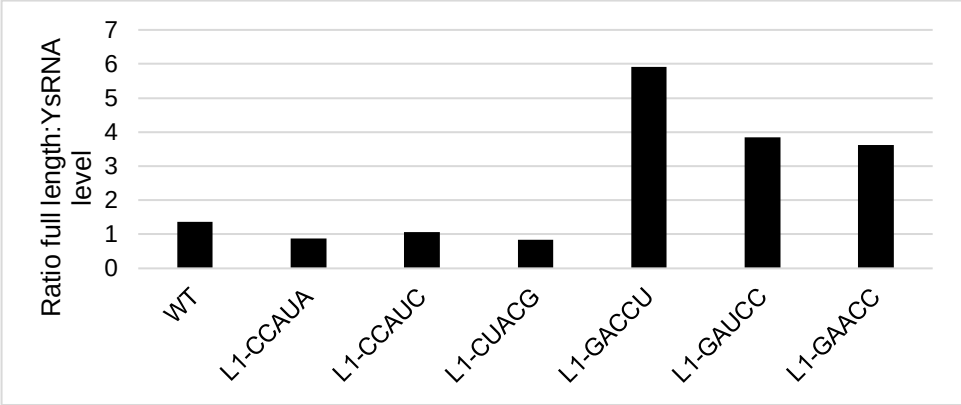
F


MCF7 + poly (I:C) 8h
Mock
hY5 WT treated
L1-CCACU
L1-CCAUC
L1-CUACG
L1-GACCU
L1-GAUCC
L1-GAACC
Full length hY5 RNA
3' end YsRNA
5' end YsRNA
U6

G


relative quantification of 3' end derived YsRNAs
1.2
1
0.8
0.6
0.4
0.2
0
WT
L1-CCAUA
L1-CCAUC
L1-CUACG
L1-GACCU
L1-GAUCC
L1-GAACC

H



I

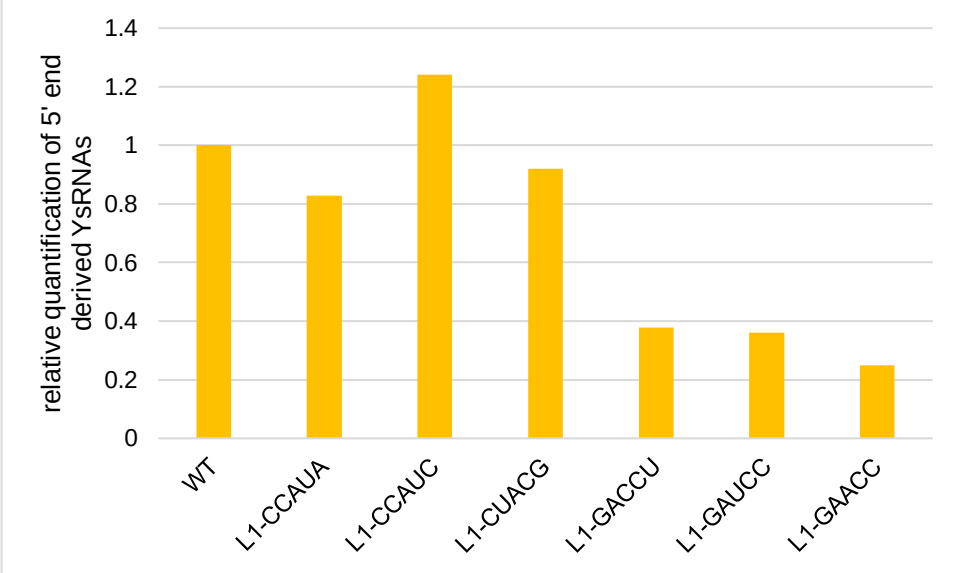


Figure 4.11. The internal loop L2b seems to be important for human Y5 RNA cleavage from both the 3' and 5' end of human Y5 RNA.

A. Predicted secondary structure of human Y5 RNA including the mutated region of loop L2b.

The secondary structure of human Y5 RNA was predicted using RNAfold and visualized using Varna. The internal loop L2b is shown in turquoise.

B. Abundances and ranks for L1 mutant hY5 RNA full length transcripts that have a high or low accumulation of YsRNAs.

The abundances were normalized across all three replicates.

C. Predicted secondary structures of mutant pool 1 hY5 RNA mutants that show high accumulation of YsRNAs compared to the full length transcript level.

The secondary structures were predicted with RNAfold under default parameters und were visualized using Varna.

D. Predicted secondary structures of mutant pool 1 hY5 RNA mutants that have a low accumulation of YsRNAs compared to the full length transcript level.

The secondary structures were predicted with RNAfold under default parameters und were visualized using Varna.

E. Analysis of secondary structure features of library 1 hY5 RNA mutants.

The table gives an overview of the nucleotide and base pair composition of stem S3, internal loop L2a and stem S3 for the tested library 1 hY5 RNA mutants.

F. Northern blot of library 1 hY5 RNA mutants that generate a high/low amount of YsRNAs compared to the full length hY5 RNA level.

Three hY5 RNA mutants that generate a high number of YsRNAs compared to the full length hY5 level were compared to the hY5 RNA mutants that generated little/no YsRNAs. Mouse 3T3 cells were transfected with library 1 hY5 RNA mutant plasmid DNAs and treated with staurosporine to induce Y RNA cleavage. Total RNA was extracted and the Northern blot membrane was probed with the 3' and 5' end probe. Total RNA extracted from MCF7 cells treated with poly (I:C) for 8 h was used as a size marker. The membrane was stripped and re-probed with U6 in order to ensure equal loading.

G. Relative quantification of 3' end derived mutant pool 1 YsRNAs.

The Northern blot of 3' end derived mutant pool 1 YsRNAs was quantified using the ImageQuant software. The optical density of each mutant was normalized to the loading control U6 and the wild type hY5 RNA derived YsRNAs. The normalized optical density of the wild type derived YsRNAs was set at 1.

H. Calculated ratio between the expression level of full length library 1 hY5 RNA mutant sequences compared to the expression level of mutant pool 1 YsRNAs.

The ratio of the normalized optical density of full length L1 hY5 RNA mutants and the 3' end derived YsRNAs was calculated and plotted on the graph.

I. Relative quantification of 3' end derived mutant pool 1 YsRNAs.

The Northern blot of 5' end derived mutant pool 1 YsRNAs was quantified using the ImageQuant software. The optical density of each mutant was normalized to the loading control U6 and the wild type hY5 RNA derived YsRNAs. The normalized optical density of the wild type derived YsRNAs was set at 1.

In further experiments I tested if the nucleotide sequence or the structure of the internal loop L2b is important for hY5 RNA cleavage. First, all possible hY5 RNA mutant combinations with the mutated RNA sequence motif UUAU were generated in-silico resulting in 256 possible sequence combinations (data not shown).

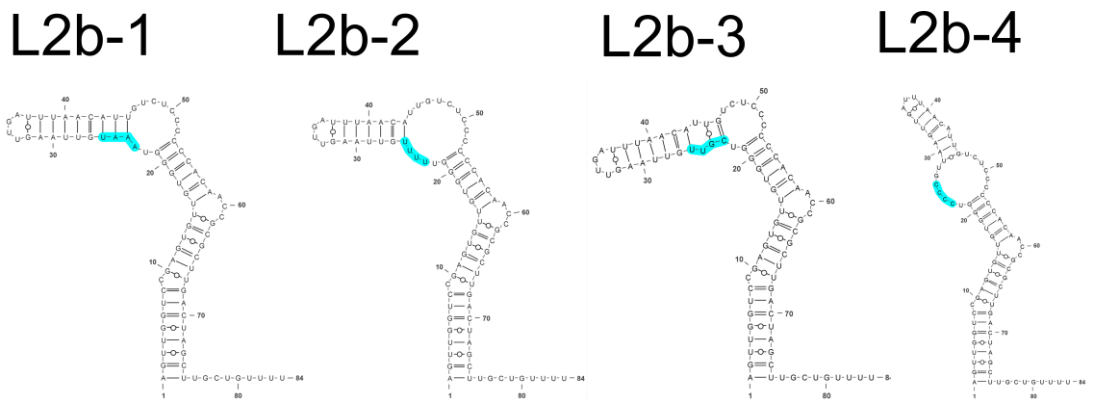
Twenty two hY5 RNA mutants out of the 256 hY5 RNA mutants folded the same as wild type hY5 RNA (**Figure 4.12, A**)

I cloned four hY5 RNA mutants out of the 22 hY5 RNA mutants that folded the same as wild type hY5 RNA individually and tested YsRNA accumulation by Northern blot. Interestingly, all four hY5 RNA mutants (AAAU, UUUU, CGUU, CCCC) with mutated internal L2b loop but same secondary structure than wild type showed decreased levels of YsRNAs from both the 5' and 3' end. Thus, these results gave some indication that the secondary structure of the internal loop L2b might not be critical for hY5 RNA cleavage and that the nucleotide sequence of this internal loop might affect both the 3' and 5' end hY5 RNA cleavage.

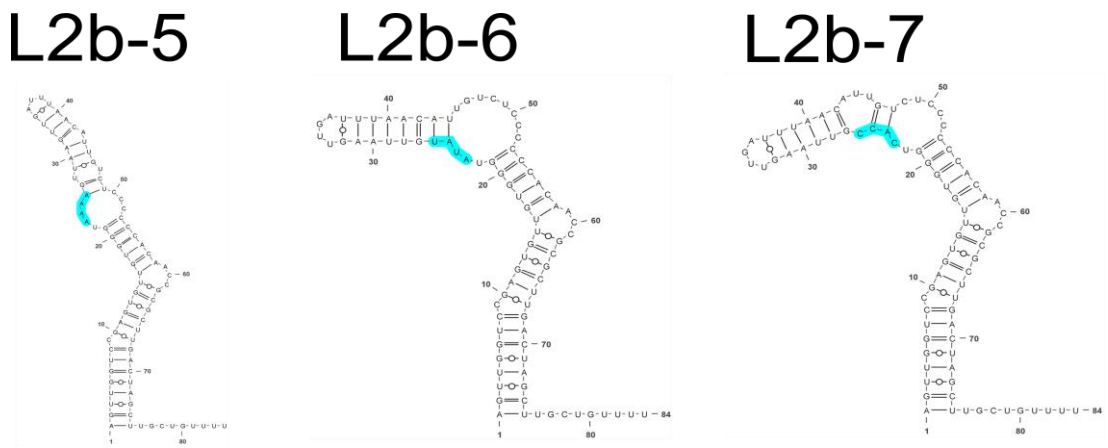
In order to further test if the internal loop L2b is critical for hY5 RNA cleavage I generated three additional hY5 RNA mutants with different secondary structure than wild type hY5 RNA. In these hY5 RNA mutants I mutated the sequence motif UUAU to AAAA, AUAU and CACC. I tested the generated hY5 RNA Loop L2b mutants experimentally by Northern blot (**Figure 4.12, C**).

Intriguingly, the Northern blot analysis showed that the mutations introduced in the hY5 RNA sequences resulted in lower accumulation of YsRNAs from both the 3' and 5' end compared to the wild type hY5 RNA (**Figure 4.12, C**)

A


L2b-1
L2b-2
L2b-3
L2b-4

B


L2b-5
L2b-6
L2b-7

C

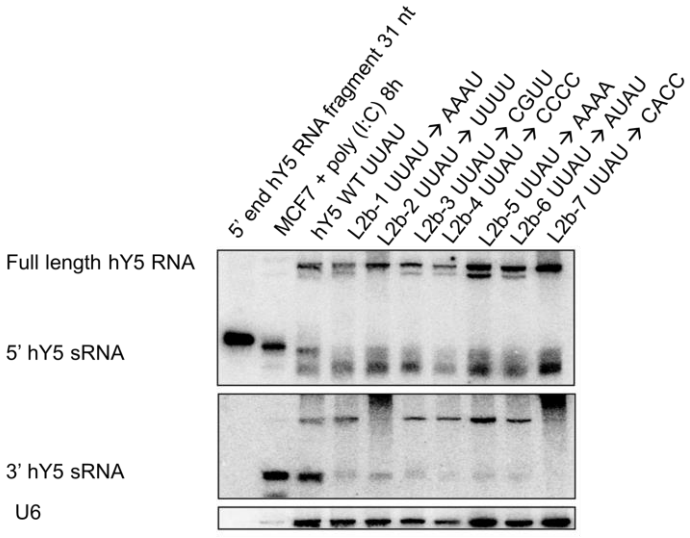

5' end hY5 RNA fragment 31 nt
MCF7 + poly (I:C) 8h
hY5 WT UUAU
L2b-1 UUAU → AAAU
L2b-2 UUAU → UUUU
L2b-3 UUAU → CGUU
L2b-4 UUAU → CCCC
L2b-5 UUAU → AAAA
L2b-6 UUAU → AUAU
L2b-7 UUAU → CACC
Full length hY5 RNA
5' hY5 sRNA
3' hY5 sRNA
U6

Figure 4.12. The wild type sequence motif UUAU of the internal loop L2b is important for 3' and 5' end hY5 RNA cleavage.

A. Predicted secondary structures of hY5 RNA mutants with mutations in the internal loop L2b that fold the same way as wild type hY5 RNA.

The mutated region is shown in turquoise. The hY5 RNA mutants have exactly the same secondary structure as wild type hY5 RNA.

B. Predicted secondary structures of hY5 RNA mutants with mutations in the internal loop L2b that have a different secondary structure to the wild type hY5 RNA.

The RNA secondary structures were predicted using RNAfold. The mutated region is shown in turquoise. The hY5 RNA mutants have different secondary structure as wild type hY5 RNA.

C. Northern blot of YsRNAs derived from hY5 RNA mutants in which the sequence of the internal loop was mutated.

The Northern blot membrane was hybridized with the 3' and 5' end hY5 probe. As a size marker a synthetic version of the 5' end derived hY5 fragment was used. For equal loading the membrane was re-probed with U6.

4.3 Discussion

With the recent advances of next generation sequencing technologies novel classes of small RNAs derived from longer transcripts have been identified. Among them, 5' end derived Y RNA fragments have been found highly enriched in extracellular vesicles released from various body fluids and have been shown to be associated with multiple types of cancer (Dhahbi et al., 2013; Dhahbi et al., 2014; Victoria Martinez et al., 2015; Tolkach et al., 2017; Tolkach et al., 2018).

Recently, Chakrabortty *et al*. demonstrated that 5' end derived hY5 fragments are functional and trigger apoptosis when primary cells were treated with EVs derived from cancer cells. Interestingly, these functional 5' end derived hY5 RNA fragments contain a conserved 8 nt motif in the upper stem of Y RNAs (5'-GUUGUGGG-3') that is also required for the initiation of chromosomal DNA replication (Gardiner et al., 2009).

A few years later further mutagenesis experiments on the conserved functional upper stem region of human Y1 were performed by Wang *et al*. in order to investigate which nucleotides or structural requirements contribute to the function of Y RNAs in chromosomal DNA replication (Wang et al., 2014). They applied biophysical methods such as solution state nuclear magnetic resonance (NMR) and far-UV circular dichroism (CD) spectrometry on mutated upper stem versions of hY1 RNA and compared the spectra to the wild type hY1 conserved stem. Sequence alignment and sequence logo analysis of vertebrate Y RNAs showed that the GUG-CAC nucleotide motif is evolutionarily conserved among all vertebrates. This indicates that this region might contribute to functionality due to selection pressure during evolution (Wang et al., 2014). When GUG was mutated to ACA the initiation of DNA replication was abolished. NMR spectroscopy and previous structure probing experiments revealed that the base pairs G17-C91 and C19-C87 are required for initiation of DNA replication. The G-C base pairs function as clamps in order to stabilize the unstable base pairs in the upper stem domain of hY1 RNA in an A-form helix. Mutations that lead to the destabilization of the A-form helical structure result in the impairment of DNA replication initiation (Wang et al., 2014).

Therefore, mutagenesis and biophysical analysis of the upper stem domain of hY1 Y RNA demonstrated that the highly evolutionarily conserved nucleotides and the stability of secondary structure in this region contributed to the function of hY1 in the initiation of DNA replication.

Interestingly, nematode sbRNAs and vertebrate Y RNAs have similar sequence and secondary structure features. Among them, the GUG-CAC motif and a further downstream penta-nucleotide motif UUAUC which were shown to be required for the initiation of DNA replication (Gardiner et al., 2009; Kowalski et al., 2015).

All the literature and findings that the 5' end derived hY5 RNA fragments are functional in combination with our results from the 3' hY5 RNA mutagenesis experiments gave us the motivation to further investigate the sequence and structural requirements for the 5' end hY5 RNA cleavage.

In this chapter, I mutated three different regions along the hY5 RNA close to the 5' end cleavage site in a high throughput mutagenesis approach using degenerate nucleotides similar to the 3' end hY5 mutagenesis approach.

The generated cDNA libraries were sent for NGS and bioinformatics analysis revealed that mutant pools 4, 5 and 6 mainly generated YsRNAs with 30 nt indicating that the hY5 RNA cleavage from the 5' end in general occurs between A30 and A31 at the position UA^A. However, the size class distribution of the YsRNA libraries also showed different sized YsRNAs with 26, 29, 30, 31, 32 or 33 nt. This heterogeneity of 5' end derived sRNAs could be explained that either the RNase involved in not very specific or that different ribonucleases are involved in Y RNA cleavage from the 5' end. Thus, it could be the case that different sizes 5' end derived YsRNAs could have different functions such as in extracellular vesicles or body fluids.

Intriguingly, Donovan *et al*. suggested that RNase L is responsible for Y RNA and tRNA cleavage during poly (I:C) treatment (Donovan et al., 2017). RNase L is a metal ion independent endoribonuclease that gets activated during viral response and cleaves single-stranded RNAs leaving a 2', 3'-cyclic phosphate. Donovan *et al*. generated small RNA cDNA libraries of stress induced RNAs containing a 2', 3'-cyclic phosphate using RtcB ligase instead of the commercial T4 RNA ligase 2 in order to capture RNAs with 2', 3'-cyclic phosphates on the 3' end of RNA fragments. Their sequencing results showed that the cleaved RNA fragments during poly (I:C) treatment all had a consensus motif UN^N at the cleavage site compared to the

untreated control samples indicating that RNase L is responsible for Y RNA cleavage. Using the RtcB ligase cDNA library preparation method for hY5 RNA the main cleavage sites were at position 26 (UU^G), 29 (UU^A) and 30 (UA^A) which is controversial because RNase L is known to cleave single-stranded RNA and hY5 RNA is double-stranded in all the cleavage sites. This would imply that the double-stranded regions would need to get unwound and refolded in a way that the RNA gets single-stranded.

These results made it even more important to determine whether it is the nucleotide sequence or the secondary structure or both which are required for human Y5 RNA cleavage to occur.

In this chapter I could clearly show by sequence logo analysis and individual substitution mutations of library 4 YsRNA mutants that at position 22 a uridine residue is critical for hY5 RNA cleavage from the 5' end.

The Northern blot analysis clearly demonstrated that when the uridine at the fifth position was mutated to adenine, cytidine, or guanine YsRNA generation from the 5' end of hY5 RNA was abolished and no YsRNAs could not be detected. Interestingly, hY5 RNA was still cleaved from the 3' end of hY5 RNAs in the substitutions hY5 RNA mutants but the YsRNA accumulation was decreased compared to the accumulation of YsRNAs derived from the wild type hY5 RNA. Additionally, in the multiple sequence alignment of all human Y RNAs I observed that this uridine residue is highly conserved among all the human Y RNAs.

Intriguingly, part of the RNA motif UGGGU that was selected for library 4 hY5 RNA mutagenesis shares the nucleotides UGGG at the position 18-21 with the consensus sequence GUAGUGGG that was previously shown to be essential for DNA replication initiation (Christov et al., 2006). Also, Chakrabortty *et al*., reported that the sequence GUUGUGGG as part of the 5' end derived hY5 RNA fragment is required for triggering apoptosis. Additionally, the results of the hY5 library 4 mutagenesis analysis confirmed that uridine at position 22 was critical for 5' end hY5 RNA cleavage and contributes to 3' end hY5 RNA cleavage. These results gave further insights that the region 5'-GUUGUGGG-3' including the uridine at position 22 might be important for functionality.

Remarkably, from all the bioinformatics, secondary structure and sequence logo analysis of hY5 library 4 and 5 hY5 RNA mutants I concluded that the internal loop

L2b at the positions 22-25 might also be involved in Y RNA cleavage from both the 3' and 5' end of hY5 RNAs.

Intriguingly, this internal loop got my interest already during the 3' end mutagenesis analysis of hY5 L1 mutants. Secondary structure prediction of hY5 L1 mutants that generated a good amount of YsRNAs compared to the full length hY5 RNA and hY5 L1 mutants that produced little/low amount of YsRNAs showed clear differences in the size of the internal loop L2b.

Individual hY5 L1 mutants that produced a good amount of YsRNAs and those that generated little amount of YsRNAs were taken for further validation.

The Northern blot analysis demonstrated that the hY5 L1 mutants with an internal loop with the sequence UUAU showed a high YsRNA accumulation compared to the hY5 L1 mutants that had a shorter internal L2b loop. From these results it could be suggested that the internal loop with the sequence UUAU affected hY5 RNA cleavage from both the 3' and 5' end.

The sequence logo analysis of the most abundant mutant pool 5 YsRNAs indicated that the sequence UAU is preferred. In order to validate if it is the nucleotide sequence UUAU or the secondary structure or both that are required for hY5 RNA cleavage to occur I tested several substitution mutants for YsRNA accumulation.

Secondary structure analysis of all the possible nucleotide combinations of the RNA motif UUAU was performed and I observed that 22 hY5 RNA mutants had the same secondary structure than the wild type hY5 RNA.

Northern blot analysis of 4 out of 22 hY5 RNA mutants with the same secondary structure but different primary sequence than wild type demonstrated that in all the hY5 RNA mutants YsRNA generation from both the 3' and 5' end of hY5 RNAs was impaired.

From these results it was suggested that both the sequence UUAU and the secondary structure of the internal loop L2b (single-stranded RNA) might contribute to hY5 RNA cleavage from both the 3' and 5' end.

Notably, in order to make this conclusion all the other 18 hY5 RNA mutants that folded the same way than wild type hY5 RNA should be tested and further investigation of mutant pool 5 hY5 RNA mutants is needed to pinpoint the exact sequence and structural requirement of this internal loop for Y RNA cleavage. In future experiments

it might be worthwhile to check if the 5' end hY5 fragment with mutations in this internal loop is still functional and triggers apoptosis.

The multiple sequence alignment of all human Y RNAs also revealed that the RNA motif UUAU is highly conserved between hY5, hY4 and hY1, whereas hY3 had one mismatch. This observed nucleotide conservation indicates that due to the selection pressure during evolution, this part of the sequence might have been important for the function of Y RNAs. Interestingly, the penta-nucleotide sequence motif UUAUC is highly conserved among nematode sbRNAs and are functional in the initiation of DNA replication. Also, it was shown in phylogenetic analysis that nematode sbRNAs are evolutionarily more related to Y5 than to the other major Y RNA clades (Kowalski et al., 2015; Duarte Junior et al., 2019).

Unexpectedly, the analysis of mutant pool 6 YsRNAs allowed us to determine the cleavage site from the 5' end which is mainly between position A30 and A31 generating YsRNAs of 30 nt and not 32 nt as previously anticipated. For this reason the analysis of the mutant pool 6 YsRNA libraries with sizes under 32 nt was difficult as it was impossible to figure out from which full length hY5 RNA mutant the sRNA was derived. The sequence logo analysis of the most abundant mutant pool 6 YsRNAs showed some preference for guanidine at all the positions. The predicted secondary structures of mutant pool 6 hY5 RNAs folded in a different way than the wild type hY5 RNA indicating that this part of the mutated region is not structure dependent and confers some structural flexibility and accessibility for RNA binding proteins and endoribonucleases responsible for hY5 RNA cleavage.

Chapter 5 Investigation of proteins involved in Y RNA cleavage

5.1 Introduction

5.1.1 Role of RNA binding proteins in cellular stress conditions

RNA binding proteins (RBPs) are key players of post-transcriptional gene regulation. They are involved in every step of RNA metabolism from transcription to processing, cellular localization, stability, transport, translation and the degradation of mRNAs (Glisovic et al., 2008; Lukong et al., 2008).

During the last decade over 1500 human RBPs have been identified and various RNA binding domains have been characterized. These domains include the most prominent RNA recognition motif (RRM), the hnRNP K homology (KH) domain, DEAD box helicase domain, double stranded RNA binding motif, Zinc finger, Pumilio domain or the Piwi/Argonaute/Zwille (PAZ) domain (Gerstberger et al., 2014).

Most of the RBPs have a modular architecture and interact with RNA via one or multiple RNA binding domains in a sequence and or structure dependent manner (Lunde et al., 2007; Dominguez et al., 2018). RBPs often form large dynamic ribonucleoprotein complexes (RNPs) with coding, non-coding or untranslated RNA (Lukong et al., 2008). The dynamic assembly and disassembly of RNPs allows the spatial and temporal regulation of gene expression at post-transcriptional level. RNPs mediate fine-tuning of the complex transcriptomic and proteomic landscape. The composition of the RNPs controls the way how cells respond and adapt to stress conditions rapidly (Beckmann et al., 2016).

The dysregulation and aberrant function of RNPs has been shown to be often associated with various different human diseases, including genetic diseases (Castello et al., 2013), neurodegenerative disorders (Cookson, 2017), neuromuscular diseases (Kolb et al., 2007), autoimmune disorders (Turner and DÍaz-Muñoz, 2018) or cancer (Pereira et al., 2017).

5.1.2 Y RNAs as part of RoRNPs

Y RNAs were initially discovered as components of the Ro60 and La ribonucleoprotein complexes in sera of patients with the autoimmune disorders systemic lupus erythematosus and Sjögren's syndrome (Lerner et al., 1981a; Wolin and Steitz, 1984). RoRNPs are involved in diverse cellular processes such as quality control pathways of various non-coding RNAs and stress response mechanisms including cell survival induced by UV irradiation or the regulation of RNA metabolism during different environmental stress conditions (Sim and Wolin, 2011).

Several studies reported that Ro60 contributes to Y RNA stability in animals and bacteria. Intriguingly, the levels of Y RNA transcripts were dramatically reduced in Ro60 -/- cells (Labbé et al., 1999) and Ro60 $^{-/-}$ mice were more sensitive to develop lupus like syndromes (Xue et al., 2003).

However, the precise biochemical function of RoRNPs and the role of Y RNAs during stress conditions remains poorly characterized.

5.1.3 Role of ribonucleases during stress

Ribonucleases (RNases) are enzymes that catalyse the cleavage of RNA at phosphodiester bonds. RNases have an essential role in the regulation of gene expression, RNA processing and turnover. During stress response RNases target mRNAs and non-coding RNAs specifically for degradation in order to respond rapidly to the change of environmental conditions. Several studies showed that during stress longer non-coding RNAs were processed into smaller RNA fragments by endonucleolytic cleavage and some of these fragments were proven to be functional (Thompson and Parker, 2009; Rother and Meister, 2011).

Rutjes et al, described that during apoptosis Y RNAs are rapidly and specifically cleaved in a caspase-dependent manner generating shorter YsRNA fragments between 22-25 nt and a larger fragment of 31 nt (Rutjes et al., 1999a). In this study it was suggested that the nucleases that cleave Y RNAs are caspase-dependent nucleases since Y RNA cleavage was similar to the cleavage of the U1-70K which was shown to be caspase-3-dependent (Casciola-Rosen et al., 1994; Rutjes et al., 1999a).

Recently, it was reported that RNase L is involved in tRNA and Y RNA cleavage during stress conditions. RNase L is an endoribonuclease that is activated in cells upon viral infection as part of the interferon induced viral response pathway. RNase L is only known to cleave single stranded RNAs and not double stranded RNA species and leaves a 3'-5' cyclic phosphate at the cleavage position.

5.1.4 Previous work and aims of this chapter

RNA interference experiments and further biochemical cleavage assays performed by a previous PhD student in our laboratory revealed that Ro60 is required for Y RNA cleavage during stress (Hall, 2013).

In order to further validate this finding and to investigate if there are other small RNAs dependent on Ro60, I constructed cDNA libraries of wild type and Ro60 $^{-/-}$ mouse embryonic stem cells (mES) untreated and treated with poly (I:C). I prepared the small RNA cDNA using High Definition (HD) adapters. The cDNA libraries that were constructed in previous work were all hugely biased for one sequence when the commercial Illumina adapter sequences were used for sRNA cDNA library construction.

The use of HD adapters has the advantage to reduce the ligation bias and to clone more different sequences compared to Illumina adapters especially when the small RNA repertoire is not large and diverse like the amount of small RNAs of mouse embryonic stem cells. It is was shown that in mouse embryonic stem cells there is only (Houbaviy et al., 2003).

The second aim of this chapter was to investigate if RNase L is involved in YsRNA generation. Recently, a study reported that RNase L is responsible for Y RNA cleavage in human cells (Donovan et al., 2017). In addition, work by a previous PhD student in our laboratory showed that RNase L might contribute to Y RNA cleavage in mouse cells upon poly (I:C) treatment (Hall, 2013). However, these results revealed that RNase L was only involved in Y RNA cleavage in mouse cells but not in human cells. In this chapter further experiments have been conducted in mouse cells and human cells to review if RNase L is required for Y RNA cleavage during stress.

5.2 Results

5.2.1 Ro60 is required for Y RNA cleavage

In order to further investigate if Ro60 is required for YsRNA generation, I generated cDNA libraries of small RNAs from wild type and Ro60 $^{-/-}$ mouse embryonic stem cells (mES) 8 hours after treated with poly (I:C) or mock treatment. The wild type and Ro60 $^{-/-}$ mES cell line were kindly gifted by Prof. Sandra Wolin. I confirmed the genotype of the wild type and Ro60 $^{-/-}$ mES cell line by a Western blot.

The Western blot showed that in the Ro60 $^{-/-}$ mES cell line no Ro60 could be detected whereas in the wild type mES cells Ro60 was expressed (**Figure 5.1**, **A**).

In order to establish whether Ro60 is involved in YsRNA generation as previously shown in our laboratory I performed a Northern blot of the extracted total RNAs from wild type and Ro60 $^{-/-}$ mES cells (**Figure 5.1**, **B**).

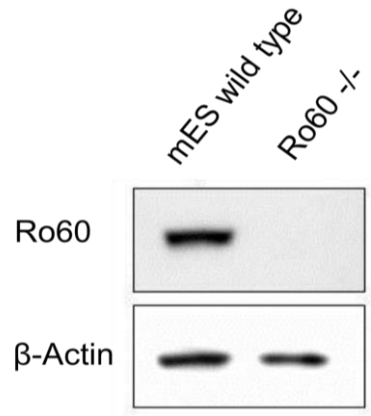
The Northern blot confirmed that Ro60 $^{-/-}$ mES cells did not generate any YsRNAs from the 3' and 5' end of mY1 when the cells were treated with poly (I:C). The wild type mES cells on the other hand showed mY1 cleavage products from both the 3' and 5' end of mY1 upon poly (I:C) treatment (**Figure 5.1**, **B**). These results showed that Ro60 is involved in the generation of Y RNA fragments in mES cells during poly (I:C) treatment.

By Adam Hall it was shown that Ro60 is also required for YsRNA generation from the 3' end of human Y5 RNA. Surprisingly, it was observed that the level of full length hY5 transcripts in the Ro60 $^{-/-}$ mES cells was not affected by the knockout of Ro60, whereas for mouse Y RNAs a dramatic decrease of full length Y RNAs could be seen (Hall, 2013).

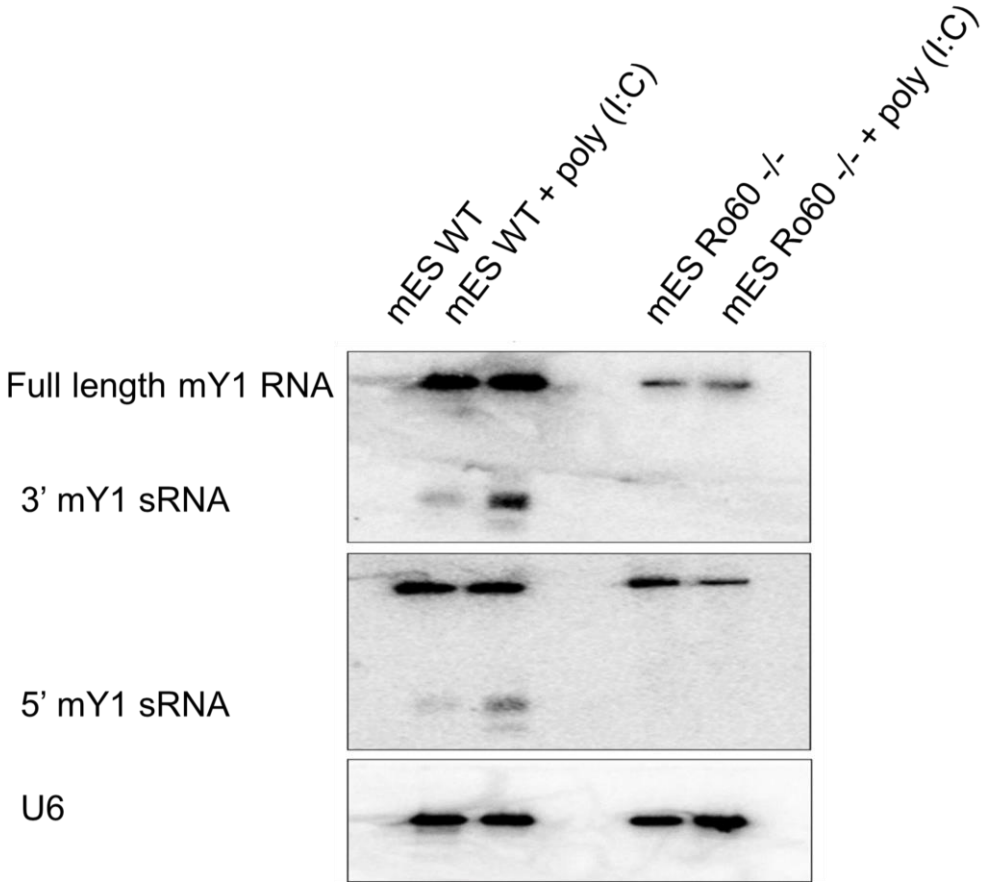
In order to confirm this finding mES wild type and Ro60 $^{-/-}$ mES cells were transfected with the hY5 wild type plasmid and treated with poly (I:C). Northern blot analysis was performed and YsRNA generation was examined.

The Northern blot experiment confirmed that Ro60 is critical for hY5 RNA cleavage from the 3' end as no YsRNA production could be observed in Ro60 $^{-/-}$ mES cells treated with poly (I:C). Additionally, it could be clearly seen that the full length hY5 RNA transcript level was not decreased in mES Ro60 $^{-/-}$ cells but the mY1 full length transcript was significantly reduced in the Ro60 $^{-/-}$ mES cells (**Figure 5.1, C**).

A


mES wild type
Ro60 -/-
Ro60
β-Actin

B


mES WT
mES WT + poly (I:C)
mES Ro60 -/-
mES Ro60 -/- + poly (I:C)
Full length mY1 RNA
3' mY1 sRNA
5' mY1 sRNA
U6

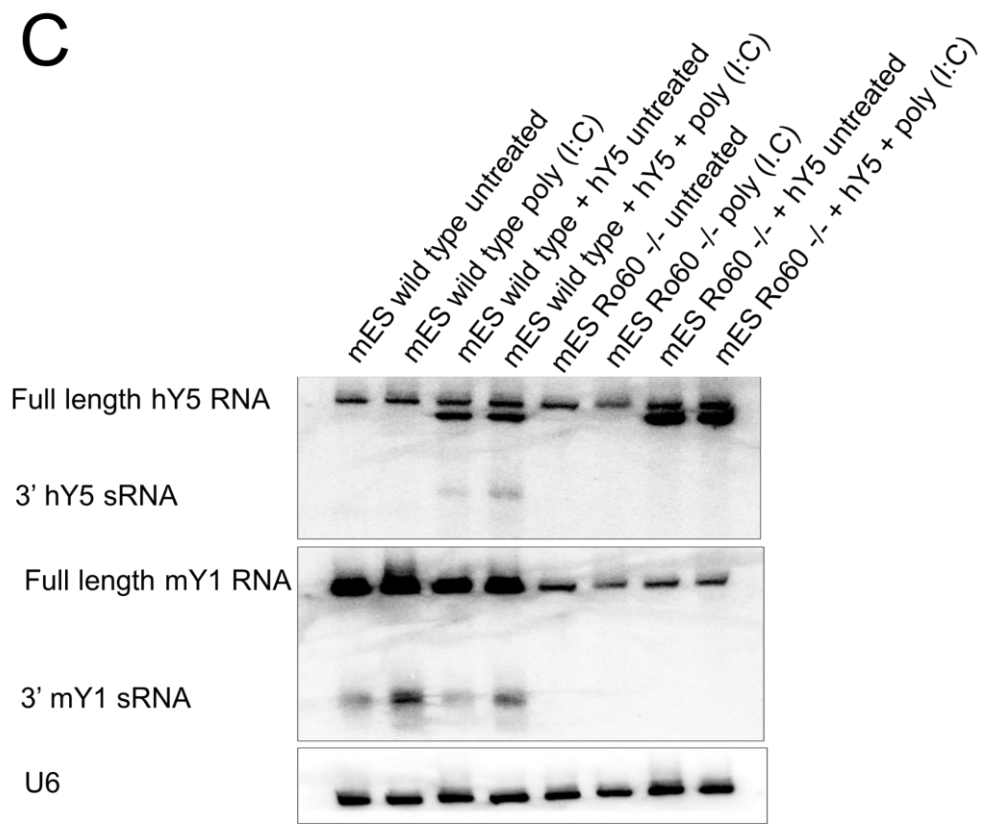


Figure 5.1. Ro60 contributes to Y RNA cleavage in mES cells.

A. Western blot of wild type and Ro60 $^{-/-}$ mES cell line.

For the detection of Ro60 an antibody against mouse Ro60 was used. As a loading control the membrane was incubated with an antibody against mouse β-actin.

B. Northern blot of mES wild type and mES Ro60 -/- cells treated and untreated with poly (I:C).

Total RNA of wild type and Ro60 $^{-/-}$ mES cells was extracted 8 hours after treated with poly (I:C) or mock treatment and subjected to Northern blot analysis. The membrane was probed with oligonucleotides complementary to the 3' or 5' end of mY1. U6 was used as a loading control.

C. Northern blot analysis of of mES wild type and mES Ro60 $^{-/-}$ cells transfected with hY5 RNA treated and untreated with poly (I:C).

Total RNA of wild type and Ro60 $^{-/-}$ mES transfected with the hY5 RNA gene and mock treated or treated with poly (I:C) was subjected to Northern blot analysis. The membrane was probed with oligonucleotides complementary to the 3' end of hY5 RNA and 3' end of mY1 RNA. As a loading control U6 was used.

As Ro60 was shown by Northern blot analysis that it is essential for YsRNA generation in human and mouse cells I further investigated by next generation sequencing of sRNA cDNA libraries of mES wild type and Ro60$^{-/-}$ cells treated and untreated with poly (I:C) if Ro60 is only responsible for YsRNA production or if the production of other sRNAs derived from longer transcripts is dependent on Ro60.

The sRNA cDNA libraries prepared by the previous PhD student Adam Hall were hugely biased for one particular sequence and it is reported that mouse embryonic stem cells have a less diverse sRNA repertoire, in this thesis I successfully modified and adapted the small RNA library construction method in order to allow efficient cloning of sRNAs.

Several recent studies reported that RNA ligases show some sequence preference in ligation of small RNAs to the adapter and that some sequences get ligated more efficiently than others (Hafner et al., 2011; Jayaprakash et al., 2011; Sorefan et al., 2012).

In 2012, our laboratory developed High definition (HD) adapters for the Illumina sequencing platform to reduce ligation bias. The HD adapters contain four degenerate nucleotides at each of the ligation ends of the commercial Illumina adapters (**Figure 5.2, A and B**).

Sorefan *et al* reported that sRNA libraries generated with HD adapters recover more different sRNA sequences and the obtained sequencing data correlated more quantitatively with the real expression level of sRNAs in the cell.

During recent years an improved method for sRNA library construction using HD adapters was developed and many libraries from several different species have been successfully generated. However, because of the low amount of sRNAs and less sRNA diversity, the protocol had to be modified and an additional gel purification step after the first ligation had to be included.

Two µg of purified total RNA was used for small RNA library preparation and the libraries were generated in three replicates. The RNA was ligated to the 3’ end HD adapter and the ligation product was separated on an 18% urea polyacrylamide gel and purified before the 5’ end HD adapter ligation, as described in detail in the materials and methods section.

In order to be able to sequence low abundance longer small RNAs with a size of 32-35 nt such as YsRNA derived fragments and to allow the differential expression analysis between Ro60 wild type and Ro60 $^{-/-}$ mouse embryonic stem cells upon poly (I:C) treatment, two cDNA library bands were purified (**Figure 5.2, C**).

A smaller cDNA library band was purified at the size of 145 nt, which included all small RNAs with sizes of 19-21 nt such as miRNAs. A longer cDNA band was also cut out between 150-160 nt as this contained all sRNAs between 32-35 nt.

For the final quantification the cDNA libraries were run on an 8% polyacrylamide gel (**Figure 5.2, D**) and all cDNA library bands were quantified using the Image Quant software and sent for next generation sequencing.

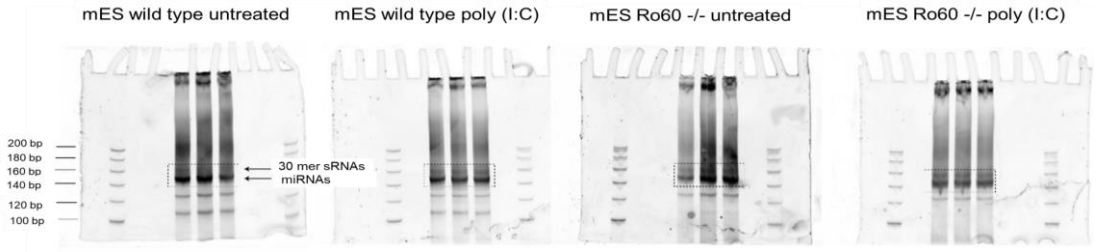
A


5' Illumina adapter
Small RNA
3' Illumina adapter
5AmMC6/GUUCAGAGUUCUACAGUCCGACGAUC
TGGAATTCTCGGGTGCCAAGG(2'3'ddC)

B


5' HD adapter
Small RNA
3' HD adapter
5AmMC6/GUUCAGAGUUCUACAGUCCGACGAUC
NNNN
NNNN
TGGAATTCTCGGGTGCCAAGG(2'3'ddC)

C


mES wild type untreated
mES wild type poly (I:C)
mES Ro60 -/- untreated
mES Ro60 -/- poly (I:C)
200 bp
180 bp
160 bp
140 bp
120 bp
100 bp
30 mer sRNAs
miRNAs

D

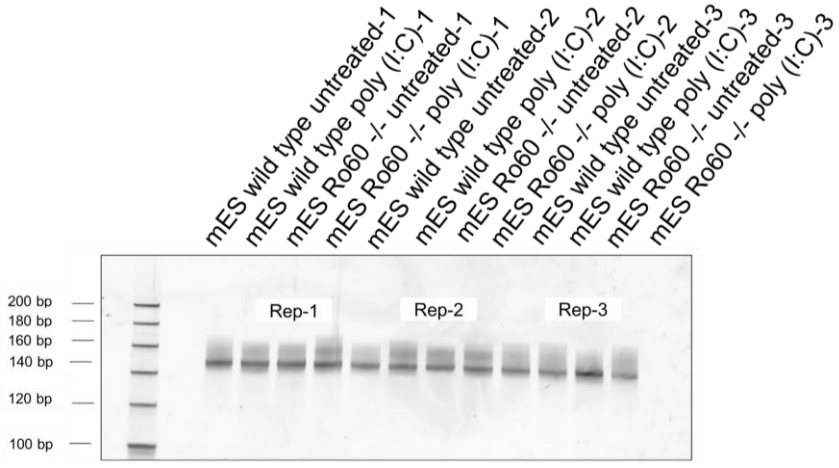

mES wild type untreated-1
mES wild type poly (I:C)-1
mES Ro60 -/- untreated-1
mES Ro60 -/- poly (I:C)-1
mES wild type untreated-2
mES wild type poly (I:C)-2
mES Ro60 -/- untreated-2
mES Ro60 -/- poly (I:C)-2
mES wild type untreated-3
mES wild type poly (I:C)-3
mES Ro60 -/- untreated-3
mES Ro60 -/- poly (I:C)-3
200 bp
180 bp
160 bp
140 bp
120 bp
100 bp
Rep-1
Rep-2
Rep-3

Figure 5.2. Generation of small RNA cDNA libraries of wild type and Ro60 $^{-/-}$ mES cells treated with poly (I:C) or mock treated using High definition adapter sequences.

A. Original Illumina adapter previously used in sRNA library construction.

During previous sRNA cDNA library construction the 5' end and 3' end of a small RNA was ligated to the original Illumina adapter sequences.

B. High definition (HD) adapter used in sRNA library preparation.

HD adapters include four degenerate nucleotides which were added to the 5' and 3' ends of the commercial Illumina adapters in order to reduce ligation bias.

C. Small RNA cDNA library construction.

Total RNA of wild type and Ro60 $^{-/-}$ mES cells was extracted after with poly (I:C) or mock treatment. The sRNA cDNA construction method using HD adapter was modified. Two µg of total RNA was used and the product after the first ligation had to be purified by gel extraction. The final PCR products were separated on a 8% PAGE and two PCR products including the miRNA band (150 nt) and a band containing 30 mer sRNAs (160 nt) were excised (shown in boxes and arrows), eluted and ethanol precipitated. A 20 bp DNA ladder was loaded as a size marker.

D. Quantification of YsRNA cDNA libraries.

One µl of each library replicate was loaded on an 8 % PAGE, alongside of a 20 bp DNA ladder and stained with SYBR Gold. The final quantification was carried out using the Image Quant software.

The bioinformatics analysis was carried out by Simon Moxon. After adapter removal, 8-13 million reads per library replicate mapping to the mouse genome (GRCm38.91) were obtained and between 70-85 % of the reads mapped to the mouse genome (see Appendix IV, **Table 0.14**).

Principal component analysis (PCA) was performed in order to check if the replicates are similar to each other, to evaluate the variation and to evaluate how well the different conditions and sRNA cDNA library replicates cluster and to visualize the variance between them (**Figure 5.3**, **A**). From the PCA plot it can be seen that one replicate of the poly (I:C) treated Ro60 $^{-/-}$ sample shown in green and a poly (I:C) treated wild type sample shown in purple are outliers. In general, the PCA plot showed that the replicates of all the different conditions cluster together and the different conditions separate nicely. For the bioinformatics analysis all the library replicates were kept.

From the size class distribution of the wild type and Ro60 $^{-/-}$ mES cells treated or not treated with poly (I:C) it can be seen (**Figure 5.3, B**) that the most abundant reads are 22 nt, which includes miRNAs. Interestingly, the size classes from 16-30 nt do not seem to be affected by the poly (I:C) treatment whereas longer sRNAs with sizes of 30-34 nt are more abundant in the poly (I:C) treated samples compared to the mock treated control.

The differential gene expression analysis using DEseq2 clearly showed by considering the p-value that sRNAs derived from mY1 and mY3 are significantly upregulated after poly (I:C) treatment (see differential expression analysis in appendix IV, **Table 0.15**). However, knockout of Ro60 resulted in a dramatic decrease of YsRNA accumulation.

For further visualization of the genomic location of the sRNAs derived from mY1 and mY3 coverage plots were generated and the genomic loci were further investigated.

A coverage plot describes the abundance of each nucleotide in all sequencing reads for a particular sequence mapped to a specific position within a reference genome. For the coverage plots of mY1 and mY3 all the reads were plotted against each position of the full length mY1/mY3 sequence.

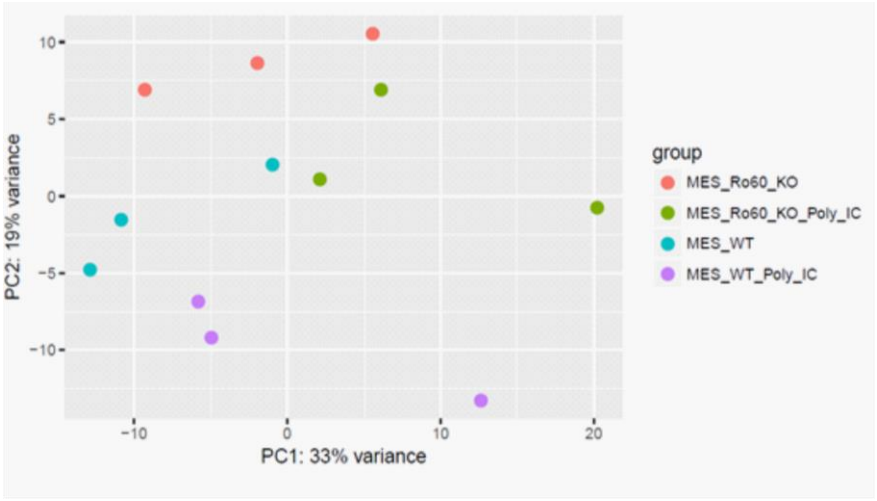
The coverage plots for mY1 and mY3 (**Figure 5.3**, **A and B**) clearly showed that the sRNAs derived from mY1 and mY3 were dramatically downregulated in the Ro60 $^{-/-}$ mES cells compared to the wild type mES and that the sRNAs were mainly generated from the 3' end of mY1 and mY3.

Interestingly, for mY3 RNA it could be observed that not only 3' and 5' end derived sRNAs were detected but also the middle part of mY3 RNA.

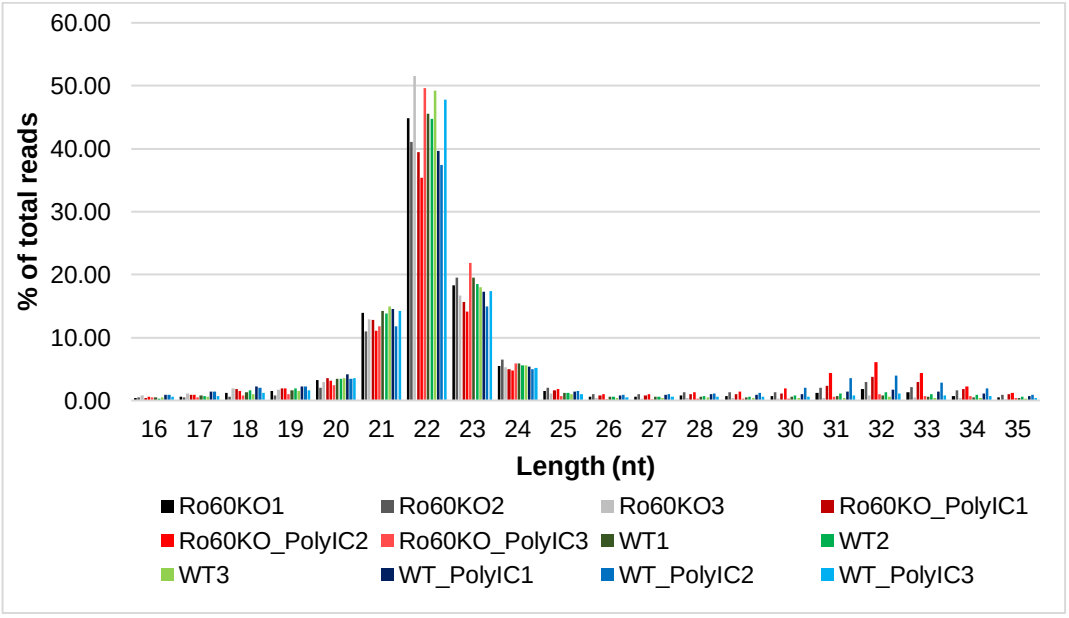
Upon treatment with poly (I:C) the sRNAs derived from mY1 and mY3 were significantly upregulated only in wild type mES cells, whereas the Ro60 $^{-/-}$ mES cells were more or less unaffected. The coverage plots clearly showed that Ro60 is required for YsRNA generation in wild type mES cells after poly (I:C) treatment. Previously, it was shown in the literature that the knockout of Ro60 had a dramatic effect on Y RNA transcript levels. It was clearly demonstrated that Ro60 stabilizes Y RNAs and that mice lacking Ro60 had drastically reduced levels of Y RNA transcripts which was reported to be around twenty times less in Ro60 $^{-/-}$ mES cells (Labbé et al., 1999; Chen et al., 2000). Thus, the absence of YsRNAs might be because the levels of Y RNA full length transcript are significantly reduced.

In order to check if the lack of YsRNAs was caused by the reduced levels of full length Y RNAs in the Ro60 $^{-/-}$ mES cells twenty times less total RNA of the wild type mES cells compared to the total RNA extracted from Ro60 $^{-/-}$ mES cells was separated on a urea polyacrylamide gel and subjected to northern blot analysis (**Figure 5.3**, **C**). The Northern blot showed that no YsRNAs could be detected in the Ro60$^{-/-}$ cells treated with poly (I:C) even when twenty times more total RNA extracted from the Ro60 $^{-/-}$ samples were loaded. Therefore, it can be concluded that Ro60 is essential for YsRNA cleavage in mouse cells.

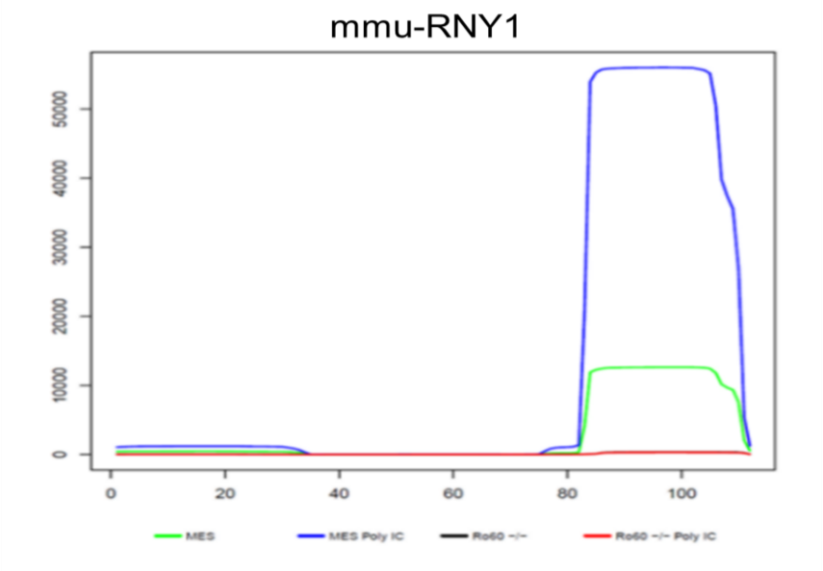
A


PC2: 19% variance
PC1: 33% variance
group
MES_Ro60_KO
MES_Ro60_KO_Poly_IC
MES_WT
MES_WT_Poly_IC

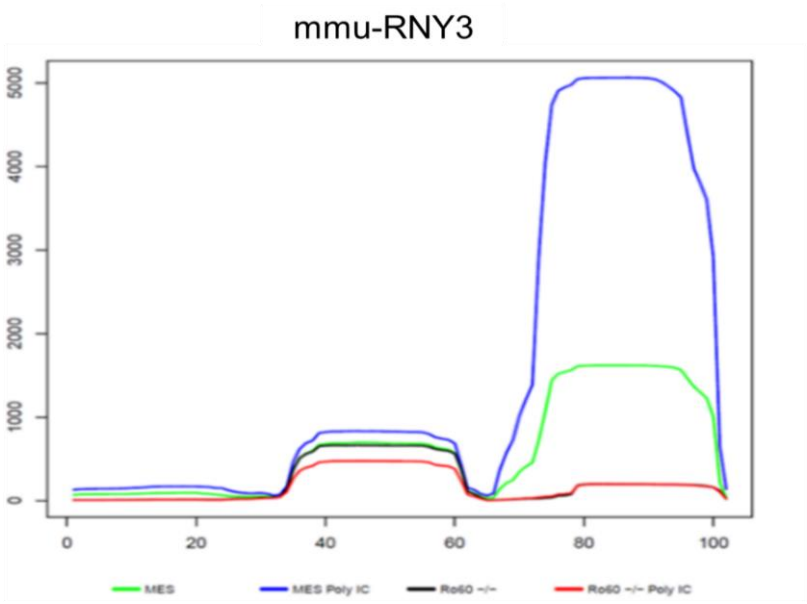
B


% of total reads
Length (nt)
Ro60KO1
Ro60KO2
Ro60KO3
Ro60KO_PolyIC1
Ro60KO_PolyIC2
Ro60KO_PolyIC3
WT1
WT2
WT3
WT_PolyIC1
WT_PolyIC2
WT_PolyIC3

C


mmu-RNY1
0
10000
20000
30000
40000
50000
0
20
40
60
80
100
MES
MES Poly IC
Ro60 -/-
Ro60 -/- Poly IC

D


mmu-RNY3
0
1000
2000
3000
4000
5000
0
20
40
60
80
100
MES
MES Poly IC
Ro60 -/-
Ro60 -/- Poly IC

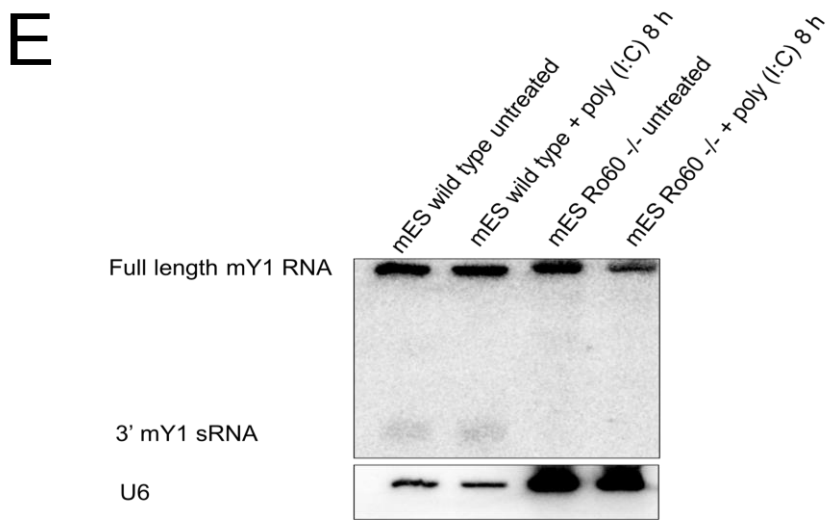


Figure 5.3. Ro60 is required for Y RNA cleavage.

A. Principal component analysis of wild type and Ro60 $^{-/-}$ mES treated or not treated with poly (I:C) sRNA cDNA libraries.

The replicates do not cluster too tight whereas the different conditions separate nicely between the wild type and Ro60 $^{-/-}$ mES cells either treated or not treated with poly (I:C).

B. Size class distribution of wild type and Ro60 $^{-/-}$ mES cells either untreated or treated with poly (I:C) sRNA cDNA libraries.

The most abundant small RNAs are at 22 nt which includes the miRNAs. For the longer fragments the most abundant sRNA peak was at 32 nt. It was observed that upon poly (I:C) treatment only the sRNAs longer than 30 nt were upregulated compared to the non-treated control samples.

C. Coverage plot and genomic location of mY1.

Upon poly (I:C) treatment the 3' end derived mY1 sRNAs were significantly upregulated in wild type mES (shown in blue) compared to the non-treated control sample (shown in green). The knockout of Ro60 resulted in a dramatic decrease of mY1 derived sRNAs (black line shows non-treated Ro60$^{-/-}$ and red line shows poly (I:C) treated Ro60$^{-/-}$ cells.

D. Coverage plot and genomic location of mY3.

The sRNAs derived from the 3' end of mY3 showed an upregulation (blue) after poly (I:C) treatment compared to the untreated sample (green). Apart from the 3' end derived mY3 sRNA also a fragment spanning the middle part of the mY3 RNA (30-60 nt) was observed. After knockout of Ro60 no sRNAs derived from mY3 could be not be detected by sequencing neither in the Ro60 $^{-/-}$ nor the Ro60 $^{-/-}$ poly (I:C) treated sample (black and red, respectively).

E. Northern blot analysis of mES wild type and mES Ro60 -/- cells treated and untreated with poly (I:C).

For Northern blot analysis 1 µg of total RNA extracted from mES wild type cells and 20 µg isolated from Ro60 $^{-/-}$ mES cells treated with poly (I:C) and mock treated were loaded. The membrane was probed with oligonucleotides complementary to the 3' or 5' end of mY1 RNA. U6 was used as a loading control.

5.2.2 Are there other Ro60 dependent small RNAs?

As Ro60 is required for YsRNA generation we wanted to further investigate if there are other small RNAs in the 30-40 nt size range apart from mY1 and mY3 sRNAs that are dependent on Ro60.

The bioinformatics analysis was performed by Simon Moxon. All the obtained reads were mapped to the mouse genome using PatMaN and a set of regions that give rise to small RNAs was defined based on mapping proximity (reads mapping within 200 bp of each other were clustered into a locus) which resulted in around 93000 different regions. All individual samples were mapped using Bowtie with the –best/-strata –m1 options (Langmead, 2010). Reads that mapped to more than one genomic position were randomly assigned to one locus. The software BEDTools was applied to count the sRNA coverage in each bam file to each locus (Quinlan, 2014). Then the read numbers were normalized and differential expression analysis was carried out by DEseq2 version 1.8.2 using default parameters (Love et al., 2014). An adjusted p-value < 0.01 was initially applied as a threshold for significant differential gene expression.

In order to explore if there are other Ro60 dependent small RNAs the sRNA I compared the expression levels between wild type and Ro60 $^{-/-}$ mES samples either treated with poly (I:C) or untreated samples. As there were no sRNAs significantly differentially expressed with a p-value adjusted < 0.01 apart from mY1, mY3 a p-value < 0.05 was considered. I loaded the annotation files of the samples into IGV in order to visualize the genomic locations where the sRNAs were derived from. I pasted the loci of the most significantly differentially expressed sRNAs (lowest p values) from the DEseq2 analysis (Appendix IV, **table 0.16**) into IGV individually and I assessed the genomic location by coverage plots whether the sRNAs were derived from intronic, exonic or non-annotated regions.

Then I tested the expression levels of a few promising potential Ro60 dependent sRNAs by Northern blot analysis in order to evaluate if the sequencing results could be validated.

The expression levels of a few promising sRNAs, including a sRNA derived from snoRNA 118, miR 296, miR 146a and miR 708 were examined by Northern blot.

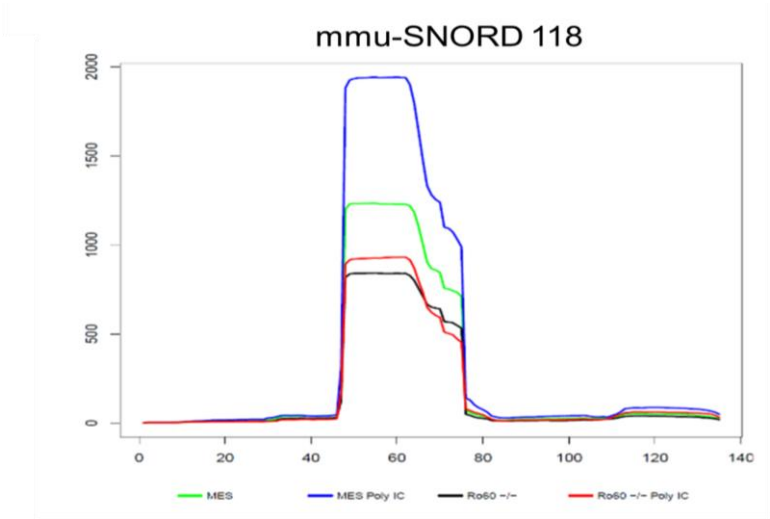
From the analysis of the coverage blots and Northern blot analysis I observed that the sRNAs derived from mY1 and mY3 were most significantly differentially expressed (Appendix IV, **table 0.15**).

The snoRNA 118 was the most promising candidate to be a Ro60 dependent small RNA among the tested sRNAs but there was just a small decrease in sRNA accumulation in Ro60 $^{-/-}$ cells treated with poly (I:C) compared to the wild type cells treated with poly (I:C) and the differential expression levels were not significant (**Figure 5.4**, **A and B**).

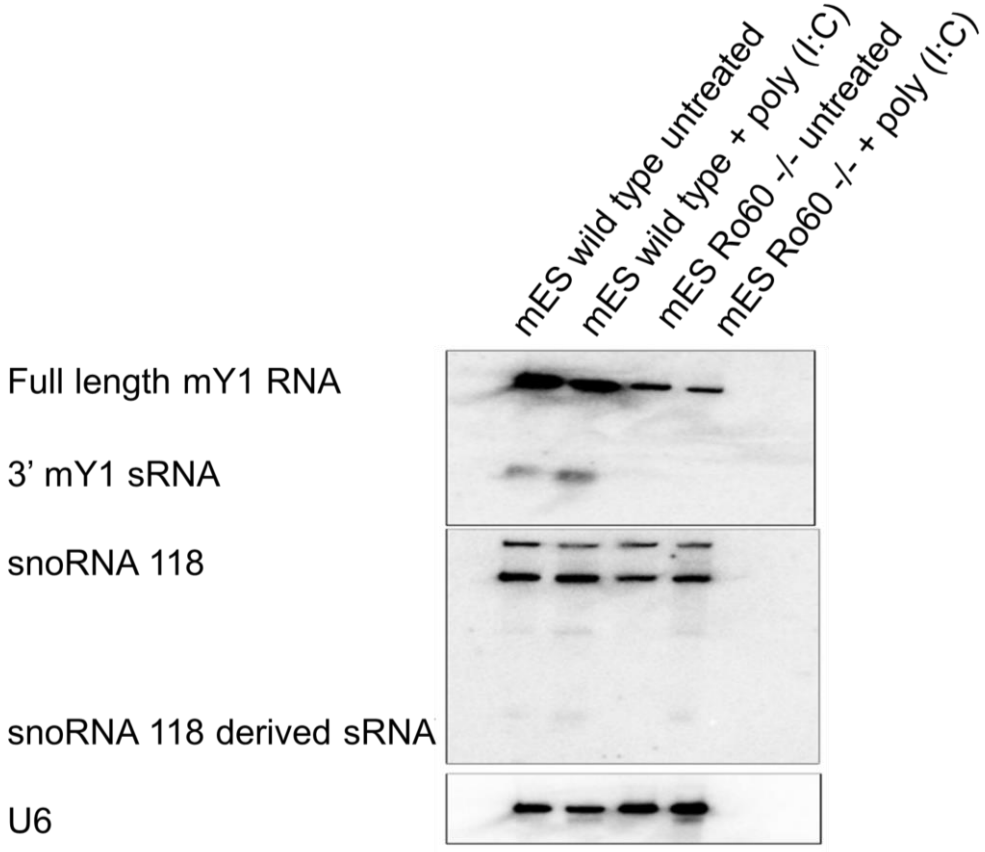
The coverage plots of the chosen and most promising miRNAs including miRNA 146a, 296 and miR 708 were generated (**Figure 5.4**, **C, D and E**) and the miRNA expression levels in mES wild type and Ro60 $^{-/-}$ cells mock treated or not treated with poly (I:C) tested by Northern blot analysis.

The Northern blot analysis of miRNA 146a, miR 296 and miR 708 indicated that Ro60 does not affect miRNA accumulation of these miRNAs (**Figure 5.4**, **E**).

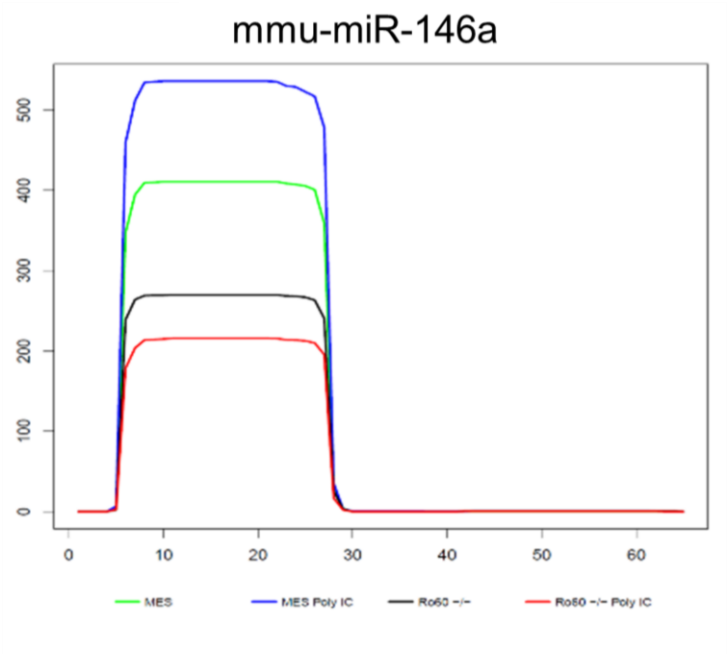
A


mmu-SNORD 118
2000
1500
1000
500
0
0
20
40
60
80
100
120
140
MES
MES Poly IC
Ro60 −/−
Ro60 −/− Poly IC

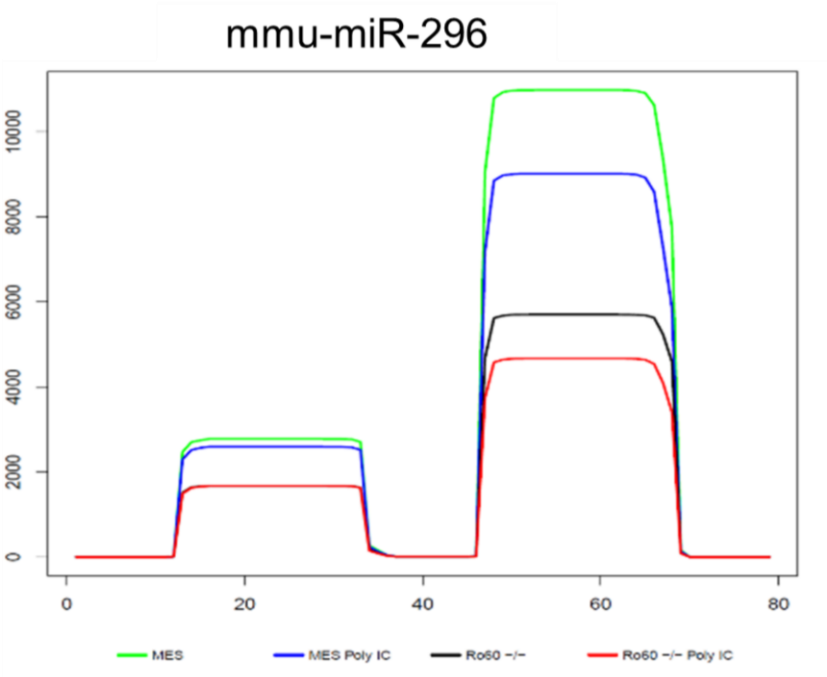
B


mES wild type untreated
mES wild type + poly (I:C)
mES Ro60 -/- untreated
mES Ro60 -/- + poly (I:C)
Full length mY1 RNA
3' mY1 sRNA
snoRNA 118
snoRNA 118 derived sRNA
U6

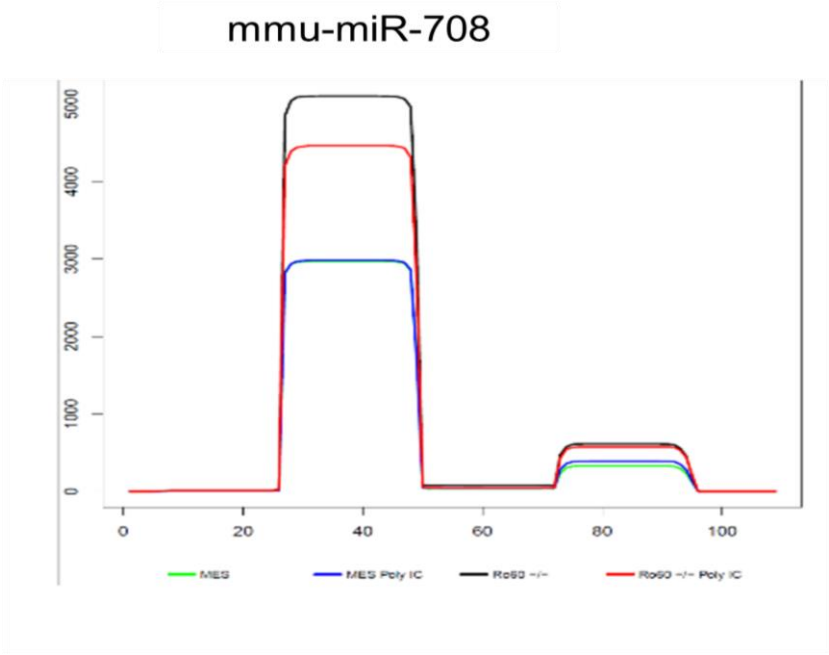
C


mmu-miR-146a
0
100
200
300
400
500
0
10
20
30
40
50
60
MES
MES Poly IC
Ro60 −/−
Ro60 −/− Poly IC

D


mmu-miR-296
0
2000
4000
6000
8000
10000
0
20
40
60
80
MES
MES Poly IC
Ro60 −/−
Ro60 −/− Poly IC

E


mmu-miR-708
0
1000
2000
3000
4000
5000
0
20
40
60
80
100
MES
MES Poly IC
Ro60 -/-
Ro60 -/- Poly IC

D

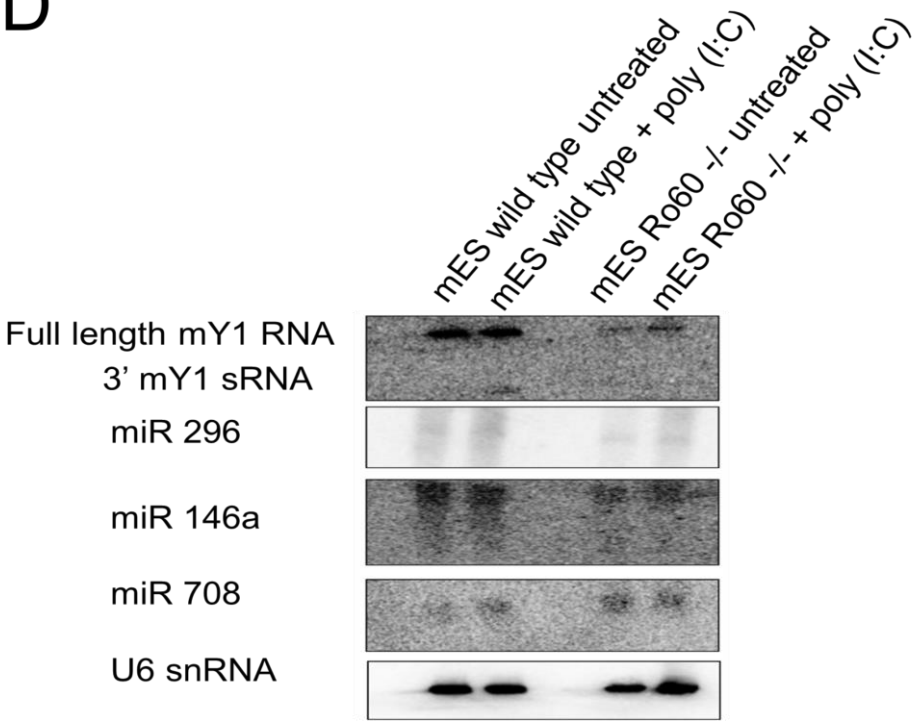

mES wild type untreated
mES wild type + poly (I:C)
mES Ro60 -/- untreated
mES Ro60 -/- + poly (I:C)
Full length mY1 RNA
3' mY1 sRNA
miR 296
miR 146a
miR 708
U6 snRNA

Figure 5.4. Ro60 seems not be involved in other Ro60 dependent sRNA generation mechanisms.

A. Coverage blot of the sRNA derived from snoRNA 118.

The snoRNA 118 derived sRNAs are upregulated upon poly (I:C) treatment compared to the untreated control. The Ro60 knockout resulted in a slight decrease of sRNA derived from snoRNA 118 compared to the wild type.

B. Northern blot analysis of the sRNA derived from snoRNA 118.

Total RNA extracted from wild type and Ro60 $^{-/-}$ cells mock treated and treated with poly (I:C) was separated on a denaturing urea polyacrylamide gel and transferred to a membrane. The membrane was probed with a probe complementary to the sRNA derived from snoRNA 118. The membrane was stripped and re-probed with U6 as a loading control.

C. Coverage plot of miR 146a.

The miR 146a showed an upregulation after poly (I:C) treatment. When Ro60 was knocked out miR 146a was down regulated. However, no difference between untreated and poly (I:C) treated wild type mES could be observed.

D. Coverage plot of miR 296.

For miR 296 no significant difference between wild type and Ro60 $^{-/-}$ mES either treated or not treated with poly (I:C) was seen.

E. Coverage plot of miR 708.

The expression levels of miR 708 in wild type and Ro60 $^{-/-}$ mES cells treated with poly (I:C) or untreated were similar.

F. Northern blot analysis of expression levels of miR146a, miR296 and miR708.

Total RNA was extracted of wild type and Ro60 $^{-/-}$ mES cells untreated and treated with poly (I:C). Northern blot analysis was performed and the membrane was probed with oligonucleotides complementary to the miRNA sequences tested. The membrane was stripped and re-probed with miR 146a, miR 296 and miR 708. As a loading control the membrane was re-probed with U6.

5.2.3 Is RNase L involved in Y RNA cleavage?

RNase L is a metal ion independent ribonuclease that is involved in the interferon induced antiviral response. The ribonuclease inhibits viral replication and cell proliferation and it is reported that it promotes apoptosis. RNase L is activated downstream of the 2' 5' oligoadenylate pathway (OAS) and double stranded RNA is the trigger for intracellular RNA cleavage. Only single stranded and not double stranded RNA is cleaved by RNase L and there is sequence specificity towards the consensus sequence UN^N.

In a recent study it was shown by Donovan *et al*. that RNase L is directly involved in the stress-induced Y RNA and tRNA cleavage (Donovan et al., 2017). Intriguingly, work by Adam Hall in our laboratory showed that RNase L contributed to Y RNA cleavage in mouse cells using wild type and RNase L $^{-/-}$ mouse fibroblast cells (mEFs) but could not be confirmed in human MCF7 cells when RNase L was silenced by a siRNA as there was no RNase L knockout available in human cells at this stage (Hall, 2013).

In this chapter I investigated if RNase L is involved in the generation of YsRNAs from the 3' and 5' end of human and mouse Y RNAs upon poly (I:C) treatment.

After I confirmed the genotype of wild type and RNase L $^{-/-}$ mEF cells the (**Figure 5.4**, **A**) the cells were either mock treated or treated with poly (I:C) in order to test if RNase L contributes to Y RNA cleavage in mEF cells. The extracted total RNA was subjected to Northern blot analysis and the membrane was probed with oligonucleotides complementary to the 3' and 5' end of mY1.

The Northern blot analysis showed that poly (I:C) treatment induced the generation of YsRNAs from the 3' and 5' end of mY1 RNA in mEF cells (**Figure 5.5, B**).

However, the knockout of RNase L in mEF cells clearly affected the accumulation of 3' and 5' end derived mY1 sRNAs and resulted in a significant decrease of mY1 sRNA levels in RNase L $^{-/-}$ cells. These results indicate that RNase L is involved in Y RNA cleavage upon poly (I:C) treatment in mEF cells.

In order to investigate if RNase L contributes to YsRNA generation in human cells, A549 lung cancer cells and A549 CRISPR RNase L $^{-/-}$ cells were either treated or not treated with poly (I:C). The extracted RNA was subjected to Northern blot analysis and Y RNA cleavage assessed (**Figure 5.4**, **C**).

The Northern blot of wild type and RNase L $^{-/-}$ A549 lung cancer cells demonstrated that the knockout of RNase L did not affect the YsRNA generation in such an extent as in the embryonic mouse fibroblast cells that were tested (**Figure 5.4**, **C and D**). After RNase L knockout I observed a downregulation of 3' and 5' hY1 and hY5 sRNAs but YsRNA production still occured (**Figure 5.5, C**).

In order to test that RNase L was completely knocked out in the cells we received I probed the Northern blot with an oligonucleotide complementary against tRNA-His and tRNA-Pro. The Northern blot showed that tRNA cleavage was completely abolished in the RNase L$^{-/-}$ A549 lung cancer cells upon poly (I:C) treatment, confirming that RNase L deletion is complete in these cells (**Figure 5.5, C**).

In order to further investigate if RNase L is required for 3' end hY5 RNA cleavage I generated hY5 substitution mutants in which the uridine of the UN^N sequence motif was replaced by adenine, guanine or cytosine. These hY5 RNA substitutions mutants were transfected into wild type and RNase L $^{-/-}$ mEF cells.

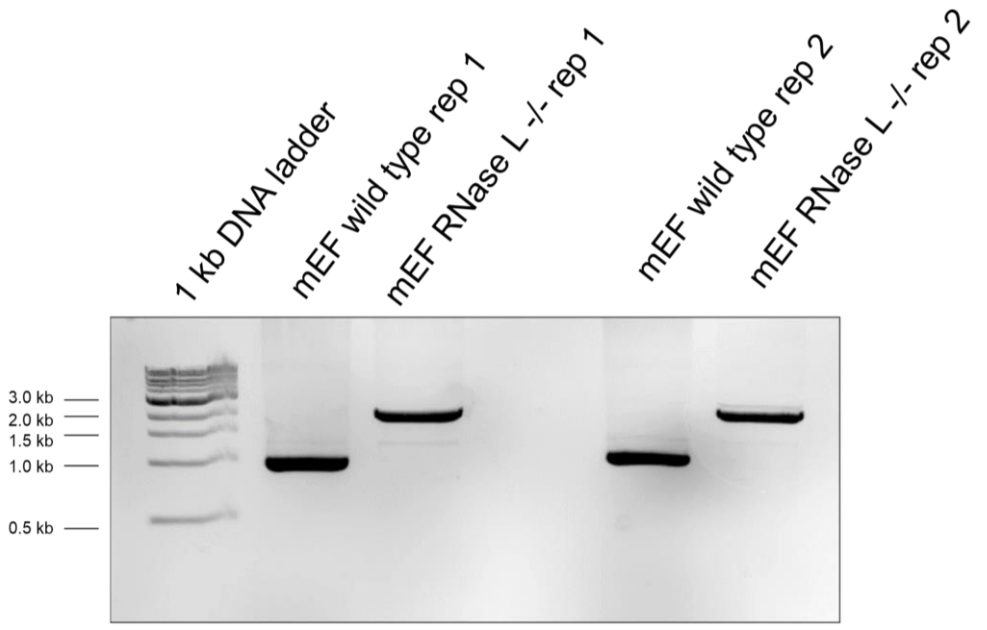
The Northern blot analysis clearly showed that the mutations introduced at position 49 did not affect the YsRNA production from the 3' or 5' end of hY5 RNAs Remarkably, it was demonstrated that the hY5 substitution mutants are expressed in RNase L $^{-/-}$ cells and cleaved into YsRNA fragments **(Figure 5.4, D**).

The Northern blot results revealed that in the hY5 substitution mutants in which the uridine in the UN^N sequence motif was replaced the hY5 RNAs still generated YsRNAs from the 3' and 5' end of hY5 RNAs. When these mutants were tested in 3T3 mouse cells) hY5 RNA cleavage could be observed from 3' and 5' end (Figure 3.11).

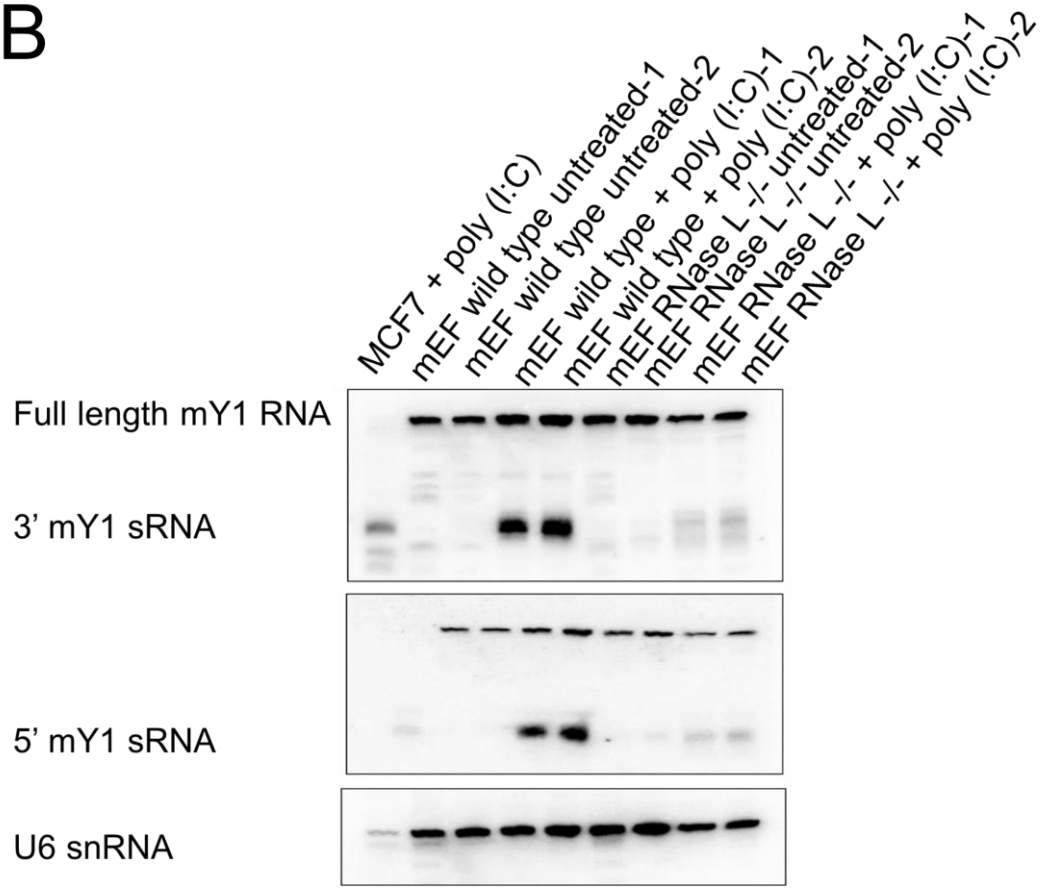
Thus, from these results it can be suggested that RNase L is not involved in hY5 RNA cleavage from the 3' end of hY5 RNAs.

However, I observed a different effect of RNase L for mouse Y1 and Y3 RNAs. In the Northern blot results I could clearly see that in RNase L $^{-/-}$ mEF cells the accumulation of YsRNAs was less compared to the wild type mEF cells. This suggested that RNase L might be involved in Y RNA cleavage from the 3' and 5' end of mouse Y RNAs but not human Y5 RNA.

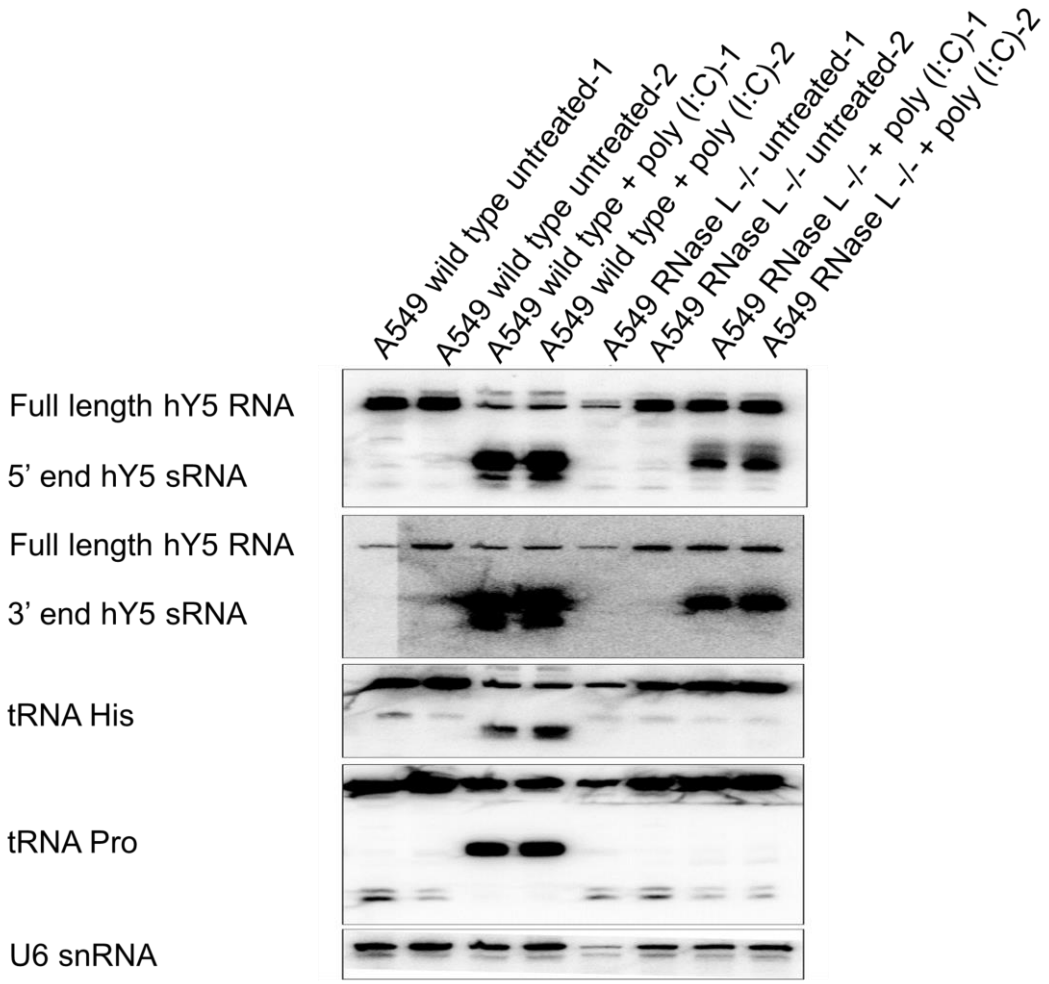
A


1 kb DNA ladder
mEF wild type rep 1
mEF RNase L -/- rep 1
mEF wild type rep 2
mEF RNase L -/- rep 2
3.0 kb
2.0 kb
1.5 kb
1.0 kb
0.5 kb

B


MCF7 + poly (I:C)
mEF wild type untreated-1
mEF wild type untreated-2
mEF wild type + poly (I:C)-1
mEF wild type + poly (I:C)-2
mEF RNase L -/- untreated-1
mEF RNase L -/- untreated-2
mEF RNase L -/- + poly (I:C)-1
mEF RNase L -/- + poly (I:C)-2
Full length mY1 RNA
3' mY1 sRNA
5' mY1 sRNA
U6 snRNA

C


A549 wild type untreated-1
A549 wild type untreated-2
A549 wild type + poly (I:C)-1
A549 wild type + poly (I:C)-2
A549 RNase L -/- untreated-1
A549 RNase L -/- untreated-2
A549 RNase L -/- + poly (I:C)-1
A549 RNase L -/- + poly (I:C)-2
Full length hY5 RNA
5' end hY5 sRNA
Full length hY5 RNA
3' end hY5 sRNA
tRNA His
tRNA Pro
U6 snRNA

D

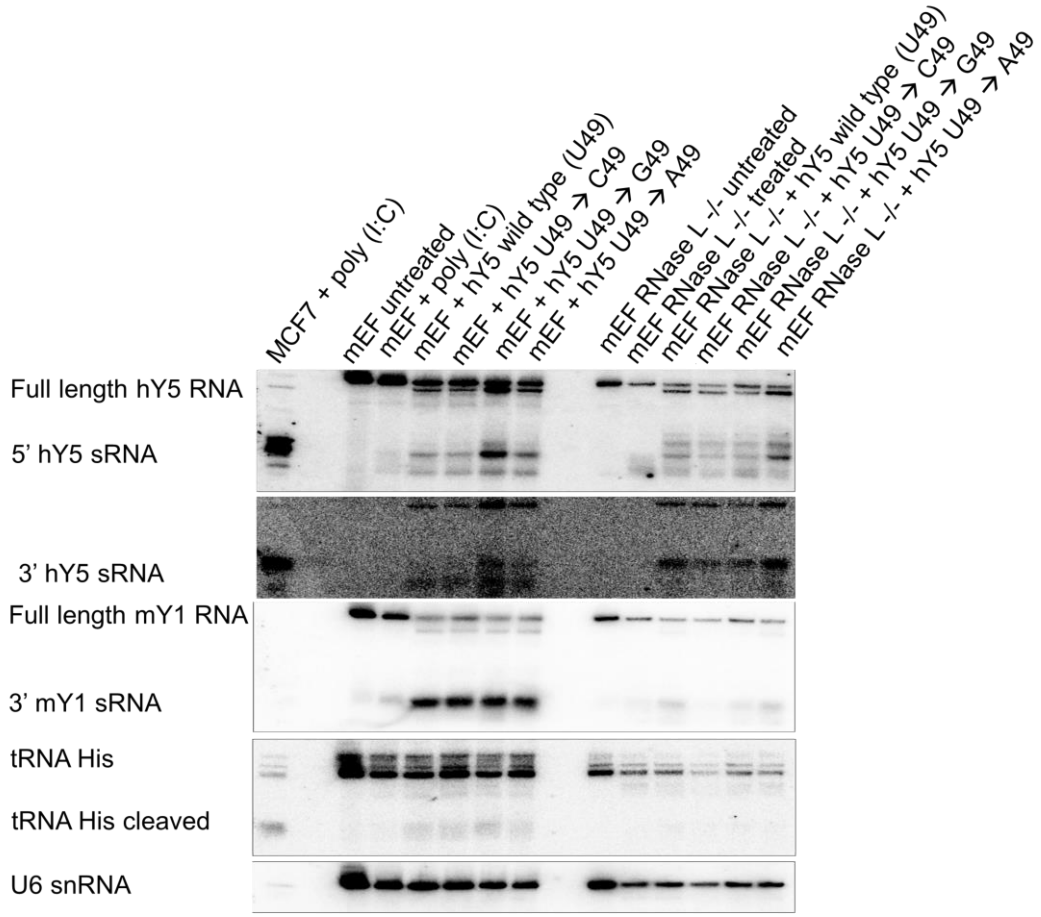


Figure 5.5. RNase L contributes to Y RNA cleavage in mEF cells and human A549 lung cancer cells upon poly (I:C) treatment but is not involved in 3' end hY5 RNA cleavage.

A. Confirmation of the RNase L $^{-/-}$ knockout mEF cell line by PCR genotyping.

Genomic DNA of wild type and RNase L $^{-/-}$ mEF cells was extracted and PCR was performed using primers flanking the insertion site of the neomycin resistance gene. The PCR products were separated on a 1% agarose gel. The size of the PCR product for the wild type is at 900 bp, whereas for the RNase L $^{-/-}$ cells it is 2.1 kb because of the neomycin gene insertion.

B. Northern blot analysis of mEF wild type and RNase L-/- cells either treated with poly (I:C) or untreated.

Total RNA of wild type and RNase L $^{-/-}$ mEF cells either treated with poly (I:C) or not treated was extracted and Northern blot analysis was performed. As a positive control total RNA extracted from MCF7 cells treated with poly (I:C) was loaded. The membrane was hybridized with oligonucleotides against the 3' and 5' end of mY1. As a loading control U6 was used.

C. Northern blot analysis of wild type and RNase L $^{-/-}$ human A549 lung cancer cells treated with poly (I:C) or untreated.

For Northern blot analysis total RNA extracted from of wild type and RNase L $^{-/-}$ human A549 lung cancer cells either treated with poly (I:C) or not treated was extracted and separated on a denaturing urea polyacrylamide gel and transferred to a membrane. As a positive control total RNA extracted from MCF7 cells treated with poly (I:C) was used. The membrane was hybridized with oligonucleotides against the 3' and 5' end of mY1. As a loading control the membrane was re-probed with U6.

D. Northern blot analysis of hY5 mutants at position 49 in mEF wild type and RNase L $^{-/-}$ cells.

Total RNA extracted from human MCF7 cells treated with poly (I:C) was used as a size marker. The hY5 substitution mutants in which the position U49 was mutated were transfected into wild type and RNase L $^{-/-}$ and treated with staurosporine for 8h. The total RNA was extracted and Northern blot was performed. The Northern blot membrane was probed with the 3' and 5' YsRNA probe as well as a probe against 3' mY1 sRNA. In order to confirm the RNase L $^{-/-}$ cell line the membrane was hybridized with a probe against tRNA-His. The blot was re-probed with U6 to demonstrate equal loading. The YsRNAs generated from the 3' end of the mutant hY5 RNAs were longer than the ones generated from wild type YsRNAs whereas the 5' end derived YsRNAs wild type had identical sizes for mutants and wild type.

5.3 Discussion

During stress Y RNAs are specifically cleaved from the 3' and 5' end. However, the biogenesis and biological significance of Y RNA processing remain obscure. In the last decades it became apparent that several classes of longer non coding RNAs including tRNAs (Lee and Collins, 2005; Thompson et al., 2008; Yamasaki et al., 2009), snoRNAs (Taft et al., 2009), snRNAs (Burroughs et al., 2011), vault RNAs (Persson et al., 2009) or rRNAs (Li et al., 2012) generate shorter sRNA fragments upon cellular stress conditions. The fact that longer non-coding RNA transcripts give rise to smaller RNA fragments showed that not only the full-length non-coding RNA transcripts but also the small RNAs derived from these transcripts can be functional.

There are several mechanisms how these sRNAs could be produced. The sRNAs can be specifically transcribed from the genomic region encoding the longer non-coding RNA transcript. Another sRNA biogenesis pathway could be that the longer non-coding RNA transcripts function as precursors that are processed by post-transcriptional endonucleolytic cleavage events.

The experiments conducted by the PhD student Adam Hall in our laboratory already indicated that Ro60 is required for Ro60. In knockdown and in vitro cleavage assays he observed that Ro60 is essential for Y RNA cleavage in human MCF7 cells (Hall, 2013).

Intriguingly, when I performed Northern blot analysis on wild type and Ro60 $^{-/-}$ mES cells either treated with poly (I:C) or mock treated I confirmed that no YsRNAs derived from mY1 and mY3 could be detected when Ro60 was knocked out.

In order to take these results further and to investigate if there are other sRNAs dependent on Ro60, I designed and generated sRNA cDNA libraries of wild type and Ro60 $^{-/-}$ mES either treated with poly (I:C) or not treated in this chapter.

For sRNA cDNA library construction I used HD adapters and successfully modified the sRNA cDNA library construction protocol. These adapters were developed in our laboratory in order to reduce the ligation bias and the sRNA library construction method has been improved (Xu et al., 2015). As mES cells contain less diverse sRNAs (Houbaviy et al., 2003) compared to other differentiated cell lines, using HD adapters during sRNA library construction might capture more diverse sRNAs compared to the Illumina adapters that were previously used. The cDNA sRNA libraries of wild type and Ro60 $^{-/-}$ mES treated with poly (I:C) and untreated control

constructed by Adam Hall in the past using the commercial Illumina adapters were hugely biased for one particular sequence (Hall, 2013).

In this chapter I successfully generated sRNA cDNA libraries of wild type and Ro60 $^{-/-}$ mES cells either treated with poly (I:C) or not treated using HD adapters and a good read coverage and percentages of reads mapping to the mouse genome were obtained.

The bioinformatics analysis and coverage plots clearly showed that Ro60 is essential for the generation of sRNAs from mY1 and mY3 RNAs. When wild type cells and Ro60 $^{-/-}$ cells were treated with poly (I:C) I could observe a dramatic downregulation of sRNAs derived from mY1 and mY3 in the Ro60 $^{-/-}$ cells. The same phenomenon I found in the Northern blot analysis when wild type and Ro60 $^{-/-}$ mES cells treated with poly (I:C) and untreated control samples were compared to each other. Intriguingly, no sRNAs generated from mY1 and mY3 could be detected by Northern blot. Remarkably, the knockout of Ro60 also had a dramatic effect on the full-length Y RNA levels of mY1 and mY3 in the Ro60 $^{-/-}$ cells.

Intriguingly, mY3 RNA also showed a sRNA fragment derived from the middle part of the mY3 RNA. This phenomenon could not be observed for the other Y RNAs in the past and it was assumed previously that the middle part of Y RNAs could not be detected due to degradation mechanisms. However, for mY3 RNA it seemed that it might have a different biogenesis mechanism and this sRNA fragment might be functional.

In conclusion, the bioinformatics analysis and Northern blot analysis confirmed that Ro60 is essential for Y RNA cleavage in mES cells upon poly (I:C) treatment. However, the role and exact mechanism on how Ro60 is involved in YsRNA cleavage remains obscure. As Ro60 is involved in the quality control of some other small non-coding RNAs, one possibility is that Ro60 might protect Y RNAs from degradation by other ribonucleases. When Ro60 is knocked out these RNases gain access and Y RNAs undergo degradation (see **Figure 5.6**). This hypothesis could also explain less accumulation of full length Y RNAs and loss of stability in Ro60 $^{-/-}$ cells compared to the wild type.

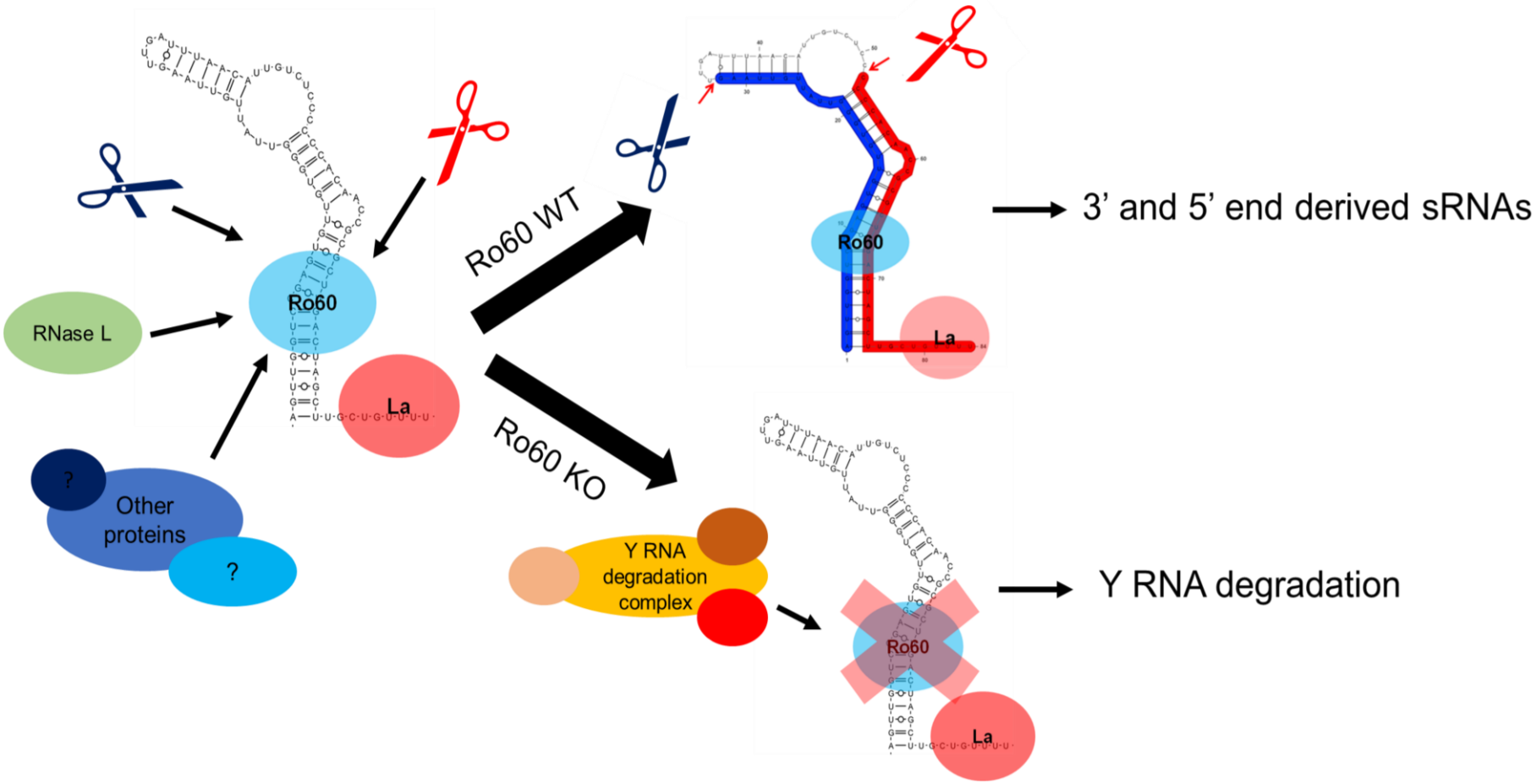

RNase L
Ro60
La
?
Other proteins
?
Ro60 WT
Ro60 KO
Ro60
La
3' and 5' end derived sRNAs
Y RNA degradation complex
Ro60
La
Y RNA degradation

Figure 5.6. Model of the role of Ro60 in YsRNA generation and protection of Y RNAs from degradation.

As it was shown that Ro60 is essential for Y RNA cleavage from the 3' end (shown in red) and 5' end (shown in blue) it was hypothesized that Ro60 is involved in the recruitment of other ribonucleases such as RNase L. From the 3' and 5' hY5 RNA mutagenesis analysis it was suggested that two different RNases might be involved in 3' and 5' end Y RNA cleavage. In this this thesis it was shown that RNase L (green) contributes to Y RNA cleavage but is not essential for Y RNA cleavage.

Therefore, Ro60 might recruit other ribonucleases/proteins involved in Y RNA cleavage (shown in blue). Interestingly, Ro60 is involved in quality control of some other small non coding RNAs and therefore Ro60 might also protect Y RNAs from degradation by preventing the access of RNases and degradation complexes (shown in orange). Thus, when Ro60 is knocked out (indicated by the red cross) full length Y RNAs undergo degradation and full length Y RNA are less stable.

In order to test if there are other Ro60 dependent sRNAs, we initially identified loci that produced higher read numbers of small RNAs in the wild type cells after poly (I:C) treatment. Then we looked at these loci and identified the ones that did not produce more small RNA reads in the Ro60 $^{-/-}$ cells after the treatment. Most of the sRNAs with the lowest p-values were further analysed regarding their genomic location and gene structure using the Integrative Genome Browser. After this preliminary analysis I further tested a few candidates by northern blot in order to check if the sequencing results can be validated.

The differential gene expression analysis of the wild type and Ro60 $^{-/-}$ mES treated with poly (I:C) and not treated showed that the sRNAs derived from mY1 and mY3 had the lowest p-values and were the most differentially expressed sRNAs.

From the analysis of the coverage plots, a sRNA derived from snoRNA 118, miR 376b, miR 296, miR 146a and miR 708 were tested by Northern blot.

However, the tested miRNAs did not seem to be dependent on Ro60 as there was no significant differential expression between wild type and Ro60 $^{-/-}$ mES samples observed in the Northern blot analysis.

Among the sRNAs tested, the snoRNA 118 derived sRNA was the most promising candidate to be dependent on Ro60. Treatment of wild type and Ro60 $^{-/-}$ mES cells with poly (I:C) resulted in an increase of sRNAs derived from snoRNA 118 compared to the control samples. However, in Ro60 $^{-/-}$ mES cells the accumulation of sRNAs generated from snoRNA 118 was only affected at a small degree was only a small decrease in Ro60$^{-/-}$ cells treated with poly (:C) compared with the sRNAs derived from mY1 and mY3.

Therefore, I concluded that Ro60 is required for Y RNA cleavage and that there seemed no other Ro60 dependent sRNA significantly differentially expressed in mES wild type and Ro60 $^{-/-}$ cells upon poly (I:C) treatment.

Apart from Ro60, I further investigated in this chapter whether RNase L contributes to YsRNA generation. In a recent study performed by Donovan *et al*. it was reported that RNase L is responsible for Y RNA cleavage at the 5' end and tRNA cleavage in human cells (Donovan et al., 2017).

Work by Adam Hall in the past demonstrated that RNase L is involved in Y RNA cleavage in mouse cells (Hall, 2013). However, the involvement of RNase L could not be confirmed in human MCF7 cells when RNase L was silenced by a siRNA.

Unfortunately, a human RNase L knockout cell line has not been generated at that time (Hall, 2013).

During my PhD I obtained new wild type and RNase L $^{-/-}$ mEF cells. After I confirmed the genotype I treated the cells with poly (I:C), performed Northern blot analysis including untreated control samples and assessed Y RNA cleavage products.

The poly (I:C) treatment resulted in an upregulation of YsRNAs from the 3' and 5' end of mY1 RNA in mEF cells compared to the untreated control. Intriguingly, the knockout of RNase L in mEF cells had a clear effect on the production of YsRNAs from both the 3' and 5' end of mY1 RNA and a reduction of YsRNAs could be observed on the Northern blot. These results indicate that RNase L is involved in Y RNA cleavage upon poly (I:C) treatment in mEF cells.

In order to check whether RNase L contributes to human Y RNA cleavage, human wild type and CRISPR RNase L $^{-/-}$ A549 cells were either treated or not treated with poly (I:C) and YsRNA accumulation tested by Northern blot. The Northern blot results demonstrated that RNase L contributes to Y RNA cleavage but YsRNA production still occurs in the RNase L $^{-/-}$ A549 cells after poly (I:C) treatment. The comparison of the Northern blot results obtained from mEF cells and human A549 lung cancer cells proposes that RNase L knockout affected the mouse cells more compared to human cells regarding YsRNA production.

I tested the complete knockout of RNase L in the human CRISPR knockout A549 cells by genotyping PCR and re-probing the Northern blot with tRNA-His and tRNA-Pro as it was reported that RNase L is responsible for tRNA cleavage upon poly (I:C) treatment (Donovan et al., 2017). In the northern blot I confirmed that both tested tRNAs were cleaved in wild type but not in RNase L $^{-/-}$ mEF cells.

The fact that the knockout of RNase L had a bigger effect on YsRNA generation in mice compared to human A549 lung cancer cells might be that other ribonucleases might be present in human cells that are involved in YsRNA production and help to maintain the cellular homeostasis when RNase L is knocked out.

Previous experiments in chapter 3 already indicated that ribonuclease(s) other than RNase L might be involved in hY5 RNA cleavage. When the uridine at position 49 which is part of the RNase L-specific UN^N sequence motif hY5 RNA cleavage still occurred from the 3' and 5' end. These results proposed that RNase L might not be involved in hY5 RNA cleavage from the 3' end. In order to validate these findings, I transfected the hY5 RNA mutants in which the uridine at position 49 was substituted

in wild type and RNase $L^{-/-}$ mEF cells, treated them with staurosporine and performed Northern blot analysis. In the Northern blot analysis I observed that YsRNAs were still generated from the 3' end of these hY5 RNA substitution mutants in RNase $L^{-/-}$ mEF cells which provided more evidence that RNase L is not required for hY5 RNA cleavage from the 3' end.

In conclusion, the obtained results suggested that apart from Ro60 and RNase L other protein complexes and ribonuclease(s) might be important for YsRNA generation that need to be investigated. One way to find the other ribonucleases critical for Y RNA cleavage could be to immunoprecipitate biotinylated Y RNAs in human and mice and perform a pulldown with streptavidin beads followed by mass spectrometry. However, this might be difficult because some protein interactions might be transient and might not be able to be capture.

Chapter 6 Summary and general discussion

Previous hY5 RNA mutagenesis experiments performed by Carly Turnbull, a previous PhD student in our laboratory indicated that hY5 RNA cleavage from the 3' end correlated with secondary structure (Turnbull, 2014). From these deletion and substitution experiments it was found that a double-stranded stem region in proximity to the internal loop in which the Y RNA cleavage occurs was critical for hY5 RNA cleavage from the 3' end.

In order to investigate sequence and structural requirements of 3' end hY5 RNA cleavage in more detail and to look at thousands of hY5 RNA mutants at the same time a high throughput mutagenesis approach of hY5 RNA close to the 3' end hY5 RNA cleavage site was performed.

In general, the aim of my thesis was to establish sequence and structural requirements of human Y5 RNA cleavage and to investigate which ribonucleases are involved in YsRNA generation.

After bioinformatics analysis I validated in chapter 3 the results from the high throughput mutagenesis approach by further secondary structure analysis, generation and testing of individual hY5 RNA mutants by Northern blot analysis.

In this chapter I could clearly show that the mutations introduced in the most abundant hY5 library 1 mutants changed the secondary structure in a way that these hY5 RNA mutants generated longer YsRNAs compared to the wild type hY5 RNA.

Interestingly, in all of these tested mutants the hY5 RNA cleavage occurred between the $2^{nd}/3^{rd}$ nt above a conserved double-stranded stem which had a high G-C content. In conclusion, in chapter 3 I was able to finally confirm that 3' end hY5 RNA cleavage correlates with secondary structure. As it is unknown which ribonuclease or proteins are involved in Y RNA cleavage these results suggest that components required for Y RNA cleavage might recognize structural features (double-stranded stem and internal loop) rather than the nucleotide sequence.

The secondary structure of hY5 RNA mutants was only computationally predicted using RNAfold and analysed by testing the expression levels of the different hY5 RNA mutants in vitro by Northern blotting. With these in vitro experiments I could already show that Y RNA cleavage correlates with the secondary structure. However, the biogenesis mechanisms of how Y RNAs are cleaved are still unknown.

In order to look in more detail into the identified secondary structure and sequence motifs in the most interesting identified hY5 RNA mutants *in vivo* structure probing of the most interesting hY5 RNA mutants identified in this thesis under stress conditions, such as poly (I:C) or staurosporine treatment would give further insights whether and/or how the secondary structure changes upon stress conditions. In these experiments it might be possible to see how the internal loops and stem domains of the hY5 RNA mutants change and it might be possible to investigate if this change in secondary structure is dynamically or not. Thus, future experiments involving *in vivo* structure analysis of all the most promising hY5 RNA mutants by SHAPE (selective 2′-hydroxyl acylation analysed by primer extension) would give further insights into the real secondary structure, interaction partners and tertiary interactions in the cellular environment. In SHAPE experiments the RNA is probed with electrophiles such as N-methylisatoic anhydride (NMIA) that react with the 2' hydroxyl group of RNA. Using primer extension the nucleotide flexibility of the RNA can be determined at single nucleotide resolution and base-paired, single stranded RNA regions or pseudoknots can be distinguished (Wilkinson et al., 2006; Hajdin et al., 2013).

Thus, for Y RNA cleavage from the 3' end we would be able to see for instance if the 5-6 nt-double-stem region and internal loop L2a are still present upon stress conditions or if the secondary structure of the hY5 RNA mutants look different *in vivo* after treatment with staurosporine.

Furthermore, *in vivo* structure probing might also be interesting for the central loop domain with the penta-nucleotide UUAUC sequence motif. Therefore, we could further investigate if this motif is present single-stranded in an internal loop, or if it is double-stranded or whether the internal loop L2b gets shortened in some of the hY5 RNA mutants. This might help to investigate the mechanism of how Y RNAs are cleaved and what are the sequence and/or structural requirements for Y RNA cleavage to occur.

In additional experiments in chapter 3 I could show that a bulge of 1 nt in the internal loop is sufficient for hY5 RNA cleavage to occur and the nucleotides at the positions 51, 52 and 53 seem not play a role in Y RNA cleavage.

As Y RNAs are cleaved from both the 3' and 5' end of hY5 RNAs and the 5' end derived hY5 RNA fragment was shown to trigger cell death in cancer cells (Chakrabortty et al., 2015) in chapter 4 I investigated if the 5' end hY5 RNA cleavage correlates with nucleotide sequence or secondary structure elements or both together.

Thus, in chapter 4, I designed and performed a high- throughput mutagenesis approach similar to the 3' end high throughput mutagenesis study.

Interestingly, the sRNA cDNA libraries of 5' end derived YsRNAs from library 4, 5 and 6 resulted in different sized YsRNAs ranging from 26-33 nt. One explanation for this heterogeneity of different sized 5' end derived YsRNAs could be that ribonucleases involved in Y RNA cleavage at the 5' end are either not very specific or different ribonucleases are responsible for different sized fragments. These different sized fragments could also have different functions for instance in extracellular vesicles or body fluids. Further it seemed, that two different ribonucleases might be involved in 3' end and 5' end hY5 RNA cleavage. The 3' end hY5 RNA cleavage correlates with secondary structure and the cleavage occurs between the 2^{nd}/3^{rd} nucleotide above a 5-6 nt double-stranded stem region, whereas the 5' end hY5 RNA seemed to the rather sequence dependent but resulting in different sized 5' end derived YsRNAs. Thus, this indicated that the Y RNA cleavage from the 5' end might not be specific and different sized fragments might fulfil different functions.

To sum up, in chapter 4 I was able to show that the RNA sequence motif of the internal loop L2b UUAU at the positions 22-25 seems to be important for both 3' and 5' end hY5 RNA cleavage. It was found that the sequence UUAU might be important, but this sequence also needs to be in a single stranded internal loop. The fact that this sequence motif affects both the 3' and 5' end Y RNA cleavage suggested that the 3' and 5' end cleavage occur dynamically but by different ribonucleases. It could be hypothesized that first one fragment gets cleaved and then other regions of the Y RNA get accessible and additional ribonucleases get recruited to the Y RNA and the Y RNA gets further processed.

Multiple sequence alignment of all human Y RNAs showed that this motif is highly evolutionarily conserved suggesting that it might be important for the function of YsRNAs. Intriguingly, mutagenesis experiments of the uridine residue at position 22 revealed that the uridine is critical for 5' end hY5 RNA cleavage and also affected cleavage at the 3' end, but to a smaller extent. Interestingly, the RNA motif 5'-GUUGUGGG-3' at the positions 14-21 which is close to the uridine residue at position 22 was in a previous study shown to be involved in triggering cell death in cancer cells (Chakrabortty et al., 2015). Additionally, this conserved motif was previously reported to be essential for the initiation of chromosomal DNA replication (Gardiner et al., 2009).

Interestingly, both nematode sbRNAs and Y RNAs have similar secondary structures with a highly conserved stem-loop secondary structure. The upper stem domain contains a evolutionarily conserved A/GUG-CAC-U nucleotide sequence motif which was reported to be crucial for the initiation of chromosomal DNA replication (Boria et al., 2010; Christov et al., 2006; Gardiner et al., 2009; Kowalski et al., 2015; Wang et al., 2014). Also, the central loop domain is conserved between both sbRNAs and Y RNAs which consists of a conserved polypyrimidine-rich penta-nucleotide motif of UUAUC at the 5' end of the loop domain (Boria et al., 2010; Wang et al., 2014; Kowalski et al., 2015).

Interestingly, the sbRNAs identified in *Bombyx mori* has a highly evolutionarily conserved sequence motif (GTGGCTTATC) with hY5 RNA and Chinese hamster chY5 RNA. This suggested that these sbRNAs may function as Y RNAs but further investigations are needed into the functions of sbRNAs in *Bombyx mori* (Duarte Junior et al., 2015; 2019).

In chapter 4 I could clearly show that uridine at position 22 of the conserved Y RNA sequence motif UGGGU is essential for 5' end hY5 RNA cleavage and might contribute to 3' end hY5 RNA cleavage. Further the internal loop with the sequence UUAU in a single-stranded RNA loop at the positions 22-25 nts affected hY5 RNA cleavage from both the 3' and 5' end. As the UUAUC penta-nucleotide sequence motif of nematode sbRNAs is essential for DNA replication and the fact that *Bombyx mori* sbRNAs are closely related to Y5 RNA it can be suggested that the function of these Y RNAs is evolutionarily conserved and Y RNA cleavage might also play a role in the initiation of DNA replication and stress conditions. Further, the high sequence conservation of the penta-nucleotide motif UUAUC between vertebrate Y RNA and sbRNAs also indicates that ribonucleases involved in Y RNA cleavage could recognize this polypyrimidine-rich penta-nucleotide motif and/or recruit other proteins or ribonucleases. In addition, it would be interesting to understand the mechanisms of how or whether the sequence, secondary structure features and/or the protein interaction partners change during stress and to look at Y RNA cleavage and DNA replication under stress conditions, such as poly (I:C) treatment. As recently the Dm1 and Dm2 sbRNAs were identified in Drosophila melanogaster it would be exciting to investigate Y RNA cleavage in insect cells in order to look if Y RNA cleavage into sRNA fragments is evolutionarily conserved.

As the penta-nucleotide motif UUAUC is also part of the 5' end derived hY5 RNA derived sRNA fragments in extracellular vesicles that trigger cell death and help tumors to establish a microenvironment there would be the need to examine the exact mechanisms of how the 5' end derived hY5 RNA fragments are generated, how they function and if they also have a role in the initiation of DNA replication also in correlation to stress and cell death. Thus, more research is needed to investigate the biogenesis and function of YsRNAs in extracellular vesicles and body fluids as Y RNAs and their fragments are highly abundant and contribute to several inflammatory diseases, including cancer (Chakrabortty et al., 2015; Hizir et al., 2017; Dhahbi et al., 2013; 2014).

In the following figure a graphical summary of establishing sequence and structural requirements of 3' and 5' end Y RNA cleavage is shown. This figure summarizes the main results that I obtained in chapter 3 and 4 in which I investigated sequence and structural requirements for human Y5 RNA from the 3' and 5' end of hY5 RNA.

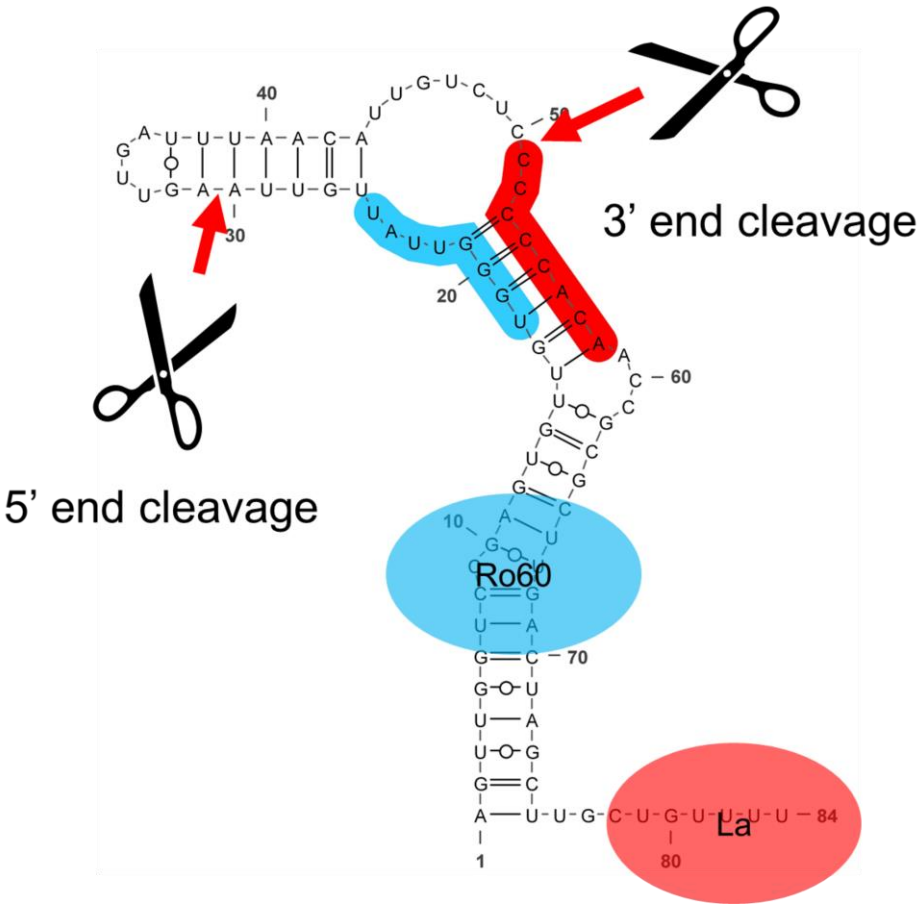


Figure 6.1. Graphical summary of the sequence and structural features identified from the 3' and 5' end high throughput mutagenesis analysis.

In chapter 3 it was shown that the 3' end hY5 RNA cleavage correlates with secondary structure. It was shown that the hY5 RNA cleavage occurred between the 2nd/3rd nt above a conserved double-stranded stem of 5-6 nts which had a high G-C content. Also, a bulge of 1 nt is sufficient for 3' end hY5 RNA cleavage.

In chapter 4 it was found that uridine at position 22 of the conserved Y RNA sequence motif UGGGU is essential for 5' end hY5 RNA cleavage and might be involved in 3' end hY5 RNA cleavage. Further the internal loop with the sequence UUAU and secondary structure (single stranded RNA) at the positions 22-25 nts affected hY5 RNA cleavage from both the 3' and 5' end. It has to be noted that the structural features identified from the 5' end high throughput mutagenesis analysis have been reported to be essential for the initiation of DNA replication in vertebrates and nematode sbRNAs.

In order to elucidate which protein(s) or ribonuclease(s) are involved in Y RNA cleavage I investigated the role of Ro60 and RNase L on Y RNA cleavage.

Previous experiments involving *in vitro* cleavage assays and knockdown experiments using RNAi conducted by Adam Hall showed that Ro60 is involved in YsRNA generation (Hall, 2013) Notably, when the Ro60 binding site was mutated, Y RNA cleavage was inhibited (Turnbull, 2014).

In order to confirm that finding and to explore if there are other sRNAs dependent on Ro60 in chapter 5 I generated cDNA libraries of wild type and Ro60 $^{-/-}$ mES cells untreated and treated with poly (I:C). In this chapter I successfully modified the cDNA library construction method and used HD adapter sequences in order to allow more diverse sequences to be cloned during sRNA library preparation.

Northern blot analysis and bioinformatics analysis clearly showed that YsRNA generation is dependent on Ro60 upon poly (I:C) treatment. Interestingly, as also previously shown by Adam Hall in our laboratory Ro60 did not affect the full-length transcript level of transfected hY5 RNA whereas the Y RNA stability of mouse Y RNAs was clearly decreased (Hall, 2013). From the literature it is known that in Ro60 $^{-/-}$ cells the Y RNA transcript stability is around twenty times less compared to the wild type (Labbé et al., 1999; Chen et al., 2000).

Further investigations will need to explore why the transfected hY5 RNA is more stable compared to endogenous mouse Y RNAs.

After differential gene expression analysis a few miRNAs that were known from the literature to be stress related were tested by Northern blot if they were Ro60 dependent. However, Northern blot analysis did not show that these miRNAs were dependent on Ro60.

One snoRNA derived sRNA (snoRNA 118) was the most promising candidate and Northern blot analysis showed a higher level in wild type mES cells treated with poly (I:C) and lower level in Ro60 $^{-/-}$ mES cells. However, the differential expression between wild type mES cells and Ro60 $^{-/-}$ cells treated and untreated with poly (I:C) was not significant. In future experiments more candidates could be tested by qRT-PCR experiments and immunoprecipitation experiments performed on the most promising candidates in order to investigate that they are bound to Ro60 protein.

To summarize, Ro60 is essential for Y RNA cleavage but seems not be involved in the generation of other Ro60 dependent sRNAs.

In addition, in chapter 5 I investigated if RNase L is responsible for Y RNA cleavage. Previous experiments performed by Adam Hall revealed that RNase L contributes to Y RNA cleavage in mouse cells, but not in human cells (Hall, 2013). Interestingly, in 2015 Donovan *et al*. claimed that RNase L is required for Y RNA cleavage. In order to investigate whether RNase L is required for Y RNA cleavage YsRNA production upon poly (I:C) treatment was tested in a mouse and human RNase L $^{-/-}$ cell line.

Northern blot analysis of wild type and RNase L $^{-/-}$ mEF cells treated with poly (I:C) or untreated indicated that RNase L is involved in Y RNA cleavage upon poly (I:C) treatment in mEF cells. However, in human A549 lung cancer and RNase L $^{-/-}$ cells it was observed that YsRNA were still generated in the RNase L $^{-/-}$ cell line.

In conclusion, RNase L contributed to Y RNA cleavage in mouse embryonic fibroblast cells and human A549 lung cancer cells but is not essential for Y RNA cleavage in human A549 lung cancer cells. Therefore, this might suggest that different ribonucleases or different repertoire of proteins are involved in YsRNA production in mouse and human cells.

Future experiments should aim to explore the RNase(s) and Y RNA interaction partners that are responsible for Y RNA cleavage in mouse and human cells. This could be achieved by immunoprecipitation of biotinylated human and mice Y RNAs followed by a pulldown using streptavidin and mass spectrometry.

References

Abbas, Y.M., Pichlmair, A., Górna, M.W., Superti-Furga, G., and Nagar, B. (2013). Structural basis for viral 5'-PPP-RNA recognition by human IFIT proteins. Nature **494,** 60-64.

Abbasi-Moheb, L., Mertel, S., Gonsior, M., Nouri-Vahid, L., Kahrizi, K., Cirak, S., Wieczorek, D., Motazacker, M.M., Esmaeeli-Nieh, S., Cremer, K., Weißmann, R., Tzschach, A., Garshasbi, M., Abedini, S.S., Najmabadi, H., Ropers, H.H., Sigrist, S.J., and Kuss, A.W. (2012). Mutations in NSUN2 cause autosomal-recessive intellectual disability. American Journal of Human Genetics **90,** 847-855.

Aftab, M.N.,He, H., Skogerbo, G., Chen, R. (2008). Microarray analysis of ncRNA expression patterns in Caenorhabditis elegans after RNAi against snoRNA associated proteins. BMC Genomics 9, 278

Alspaugh, M.A., and Tan, E.M. (1975). Antibodies to cellular antigens in Sjögren's syndrome. Journal of Clinical Investigation **55,** 1067-1073.

Amort, M., Nachbauer, B., Tuzlak, S., Kieser, A., Schepers, A., Villunger, A., and Polacek, N. (2015). Expression of the vault RNA protects cells from undergoing apoptosis. Nature Communications **6,** 7030.

Armour, C.D., Castle, J.C., Chen, R., Babak, T., Loerch, P., Jackson, S., Shah, J.K., Dey, J., Rohl, C.A., Johnson, J.M., and Raymond, C.K. (2009). Digital transcriptome profiling using selective hexamer priming for cDNA synthesis. Nature Methods **6,** 647-649.

Balakin, A.G., Smith, L., and Fournier, M.J. (1996). The RNA World of the Nucleolus: Two Major Families of Small RNAs Defined by Different Box Elements with Related Functions. Cell **86,** 823-834.

Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell **116,** 281-297.

Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. Cell **136,** 215-233.

Beckmann, B.M., Castello, A., and Medenbach, J. (2016). The expanding universe of ribonucleoproteins: of novel RNA-binding proteins and unconventional interactions. Pflügers Archiv - European Journal of Physiology **468,** 1029-1040.

Belisova, A., Semrad, K., Mayer, O., Kocian, G., Waigmann, E., Schroeder, R., and Steiner, G. (2005). RNA chaperone activity of protein components of human Ro RNPs. RNA **11,** 1084-1094.

Bell S.P. and Dutta A.(2002). DNA replication in eukaryotic cells. Annual Review of Biochemistry **71**, 333–374.

Ben-Chetrit, E., Chan, E.K., Sullivan, K.F., and Tan, E.M. (1988). A 52-kD protein is a novel component of the SS-A/Ro antigenic particle. The Journal of Experimental Medicine **167,** 1560-1571.

Berger, W., Steiner, E., Grusch, M., Elbling, L., and Micksche, M. (2008). Vaults and the major vault protein: Novel roles in signal pathway regulation and immunity. Cellular and Molecular Life Sciences **66,** 43.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature **409,** 363-366.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R.*, et al.* (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature **447,** 799-816.

Boccitto, M., and Wolin, S.L. (2019). Ro60 and Y RNAs: structure, functions, and roles in autoimmunity, Critical Reviews in Biochemistry and Molecular Biology **54:2**, 133-152

Bohnsack, M.T., Czaplinski, K., and Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. RNA **10,** 185-191.

Boire, G., and Craft, J. (1989). Biochemical and immunological heterogeneity of the Ro ribonucleoprotein particles. Analysis with sera specific for the Ro(hY5) particle. Journal of Clinical Investigation **84,** 270-279.

Boire, G., Gendron, M., Monast, N., Bastin, B., and Ménard, H.A. (1995). Purification of antigenically intact Ro ribonucleoproteins; biochemical and immunological evidence that the 52-kD protein is not a Ro protein. Clinical and Experimental Immunology **100,** 489-498.

Boria, I., Gruber, A.R., Tanzer, A., Bernhart, S.H., Lorenz, R., Mueller, M.M., Hofacker, I.L., and Stadler, P.F. (2010). Nematode sbRNAs: Homologs of Vertebrate Y RNAs. Journal of Molecular Evolution **70,** 346-358.

Bouffard, P., Barbar, E., Briere, F., and Boire, G. (2000). Interaction cloning and characterization of RoBPI, a novel protein binding to human Ro ribonucleoproteins. RNA **6,** 66-78.

Boulanger, C., Chabot, B., Ménard, H.A., and Boire, G. (1995). Autoantibodies in human anti-Ro sera specifically recognize deproteinized hY5 Ro RNA. Clinical and Experimental Immunology **99,** 29-36.

Brameier, M., Herwig, A., Reinhardt, R., Walter, L., and Gruber, J. (2011). Human box C/D snoRNAs with miRNA like functions: expanding the range of regulatory RNAs. Nucleic Acids Research **39,** 675-686.

Burroughs, A.M., Ando, Y., de Hoon, M.L., Tomaru, Y., Suzuki, H., Hayashizaki, Y., and Daub, C.O. (2011). Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. RNA Biology **8,** 158-177.

Buzas, E.I., György, B., Nagy, G., Falus, A., and Gay, S. (2014). Emerging role of extracellular vesicles in inflammatory diseases. Nature Reviews Rheumatology **10,** 356-364.

Cambier, L., de Couto, G., Ibrahim, A., Echavez, A.K., Valle, J., Liu, W., Kreke, M., Smith, R.R., Marbán, L., and Marbán, E. (2017). Y RNA fragment in extracellular vesicles confers cardioprotection via modulation of IL-10 expression and secretion. EMBO Molecular Medicine **9,** 337-352.

Casciola-Rosen, L.A., Anhalt, G., and Rosen, A. (1994). Autoantigens targeted in systemic lupus erythematosus are clustered in two populations of surface structures on apoptotic keratinocytes. The Journal of Experimental Medicine **179,** 1317-1330.

Castello, A., Fischer, B., Hentze, M.W., and Preiss, T. (2013). RNA-binding proteins in Mendelian disease. Trends in Genetics **29,** 318-327.

Chakrabarti, A., Banerjee, S., Franchi, L., Loo, Y.-M., Gale, M., Núñez, G., and Silverman, Robert H. (2015). RNase L Activates the NLRP3 Inflammasome during Viral Infections. Cell Host & Microbe **17,** 466-477.

Chakrabortty, S.K., Prakash, A., Nechooshtan, G., Hearn, S., and Gingeras, T.R. (2015). Extracellular vesicle-mediated transfer of processed and functional RNY5 RNA. RNA **21,** 1966-1979.

Chan, S., Choi, E.-A., and Shi, Y. (2011). Pre-mRNA 3'-end processing complex assembly and function. Wiley interdisciplinary reviews. RNA **2,** 321-335.

Chen, C.-J., and Heard, E. (2013). Small RNAs derived from structural non-coding RNAs. Methods **63,** 76-84.

Chen, X., Quinn, A.M., and Wolin, S.L. (2000). Ro ribonucleoproteins contribute to the resistance of Deinococcus radiodurans to ultraviolet irradiation. Genes and Development **14,** 777-782.

Chen, X., Sim, S., Wurtmann, E.J., Feke, A., and Wolin, S.L. (2014). Bacterial noncoding Y RNAs are widespread and mimic tRNAs. Rna **20,** 1715-1724.

Chen, X., Smith, J.D., Shi, H., Yang, D.D., Flavell, R.A., and Wolin, S.L. (2003). The Ro Autoantigen Binds Misfolded U2 Small Nuclear RNAs and Assists Mammalian Cell Survival after UV Irradiation. Current Biology **13,** 2206-2211.

Chen, X., Wurtmann, E.J., Van Batavia, J., Zybailov, B., Washburn, M.P., and Wolin, S.L. (2007). An ortholog of the Ro autoantigen functions in 23S rRNA maturation in D. radiodurans. Genes and Development **21,** 1328-1339.

Chen, X., Taylor, D.W., Fowler, C.C., Galan, J.E., Wang, H.W., and Wolin, S.L. (2013). An RNA degradation machine sculpted by Ro autoantigen and noncoding RNA. Cell **153,** 166-177.

Chiu, Y.-L., Witkowska, H.E., Hall, S.C., Santiago, M., Soros, V.B., Esnault, C., Heidmann, T., and Greene, W.C. (2006). High-molecular-mass APOBEC3G complexes restrict Alu retrotransposition. Proceedings of the National Academy of Sciences **103,** 15588-15593.

Chow, R.D., and Chen, S. (2018). Sno-derived RNAs are prevalent molecular markers of cancer immunity. Oncogene.

Christov, C.P., Trivier, E., and Krude, T. (2008). Noncoding human Y RNAs are overexpressed in tumours and required for cell proliferation. British Journal of Cancer **98,** 981-988.

Christov, C.P., Gardiner, T.J., Szüts, D., and Krude, T. (2006). Functional requirement of noncoding Y RNAs for human chromosomal DNA replication. Molecular and Cellular Biology **26,** 6993-7004.

Christov, C.P., Dingwell, K.S., Skehel, M., Wilkes, H.S., Sale, J.E., Smith, J.C., and Krude, T. (2018). A NuRD Complex from Xenopus laevis Eggs Is Essential for DNA Replication during Early Embryogenesis. Cell Reports **22,** 2265-2278.

Cole, C., Sobala, A., Lu, C., Thatcher, S.R., Bowman, A., Brown, J.W.S., Green, P.J., Barton, G.J., and Hutvagner, G. (2009). Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. RNA **15,** 2147-2160.

Coleman, B.M., and Hill, A.F. (2015). Extracellular vesicles--Their role in the packaging and spread of misfolded proteins associated with neurodegenerative diseases. Seminars in Cell & Developmental biology **40,** 89-96.

Collart, C., Christov, C.P., Smith, J.C., and Krude, T. (2011). The midblastula transition defines the onset of Y RNA-dependent DNA replication in Xenopus laevis. Molecular and Cellular Biology **31,** 3857-3870.

Colombo, M., Raposo, G., and Théry, C. (2014). Biogenesis, Secretion, and Intercellular Interactions of Exosomes and Other Extracellular Vesicles. Annual Review of Cell and Developmental Biology **30,** 255-289.

Cookson, M.R. (2017). RNA-binding proteins implicated in neurodegenerative diseases. Wiley Interdisciplinary Reviews: RNA **8**.

Cooper, D.A., Jha, B.K., Silverman, R.H., Hesselberth, J.R., and Barton, D.J. (2014). Ribonuclease L and metal-ion-independent endoribonuclease cleavage sites in host and viral RNAs. Nucleic Acids Research **42,** 5202-5216.

Crick, F. (1970). Central Dogma of Molecular Biology. Nature **227,** 561-563.

Crick, F.H.C. (1958). On Protein Synthesis. The Symposia of the Society for Experimental Biology **12,** 138-163.

Crooks, G.E., Hon, G., Chandonia, J.M., and Brenner, S.E. (2004). WebLogo: A sequence logo generator. Genome research **14,** 1188-1190.

Cui, L., Guo, X., Qi, Y., Qi, X., Ge, Y., Shi, Z., Wu, T., Shan, J., Shan, Y., Zhu, Z., and Wang, H. (2010). Identification of microRNAs involved in the host response to enterovirus 71 infection by a deep sequencing approach. Journal of biomedicine & biotechnology **2010,** 425939.

Darty, K., Denise, A., and Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. Bioinformatics **25,** 1974-1975.

Deng, W., Zhu, X., Skogerbo, G., Zhao, Y., Fu, Z., Wang, Y., He, H., Cai, L., Sun, H., Liu, C., Li, B., Bai, B., Wang, J., Jia, D., Sun, S., He, H., Cui, Y., Wang, Y., Bu, D., and Chen, R. (2006). Organization of the Caenorhabditis elegans small non-coding transcriptome: genomic features, biogenesis, and expression. Genome Research **16,** 20-29.

Denli, A.M., Tops, B.B.J., Plasterk, R.H.A., Ketting, R.F., and Hannon, G.J. (2004). Processing of primary microRNAs by the Microprocessor complex. Nature **432,** 231-235.

Deutscher, S.L., Harley, J.B., and Keene, J.D. (1988). Molecular analysis of the 60-kDa human Ro ribonucleoprotein. Proceedings of the National Academy of Sciences of the United States of America **85,** 9479-9483.

Dhahbi, J.M., Spindler, S.R., Atamna, H., Boffelli, D., and Martin, D.I. (2014). Deep Sequencing of Serum Small RNAs Identifies Patterns of 5' tRNA Half and Y RNA Fragment Expression Associated with Breast Cancer. Biomarkers in cancer **6,** 37-47.

Dhahbi, J.M., Spindler, S.R., Atamna, H., Boffelli, D., Mote, P., and Martin, D.I. (2013). 5'-Y RNA fragments derived by processing of transcripts from specific Y RNA genes and pseudogenes are abundant in human serum and plasma. Physiological Genomics **45,** 990-998.

Dominguez, D., Freese, P., Alexis, M.S., Su, A., Hochman, M., Palden, T., Bazile, C., Lambert, N.J., Van Nostrand, E.L., Pratt, G.A., Yeo, G.W., Graveley, B.R., and Burge, C.B. (2018). Sequence, Structure, and Context Preferences of Human RNA Binding Proteins. Molecular Cell **70,** 854-867.e859.

Dong, B., and Silverman, R.H. (1995). 2-5A-dependent RNase Molecules Dimerize during Activation by 2-5A. Journal of Biological Chemistry **270,** 4133-4137.

Donovan, J., Rath, S., Kolet-Mandrikov, D., and Korennykh, A. (2017). Rapid RNase L–driven arrest of protein synthesis in the dsRNA response without degradation of translation machinery. RNA **23,** 1660-1671.

Duarte Junior, F.F., Lima Neto, Q.A.D., Rando, F.D.S., Freitas, D.V.B.D., Pattaro Júnior, J.R., Polizelli, L.G., Munhoz, R.E.F., Seixas, F.A.V., and Fernandez, M.A. (2015). Identification and molecular structure analysis of a new noncoding RNA, a sbRNA homolog, in the silkworm Bombyx mori genome. Molecular BioSystems **11,** 801-808.

Duarte Junior, F.F., Bueno, S.A.A., Pedersen, S.L., Rando, F.D.S., Pattaro Junior, J.R. et al. (2019). Identification and characterization of stem-bulge RNAs in Drosophila melanogaster. RNA biology **16**:3, 330-339

Elbashir, S.M., Lendeckel, W., and Tuschl, T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. Genes and Development **15,** 188-200.

Emara, M.M., Ivanov, P., Hickman, T., Dawra, N., Tisdale, S., Kedersha, N., Hu, G.-F., and Anderson, P. (2010). Angiogenin-induced tRNA-derived Stress-induced RNAs Promote Stress-induced Stress Granule Assembly. The Journal of Biological Chemistry **285,** 10959-10968.

Ender, C., Krek, A., Friedländer, M.R., Beitzinger, M., Weinmann, L., Chen, W., Pfeffer, S., Rajewsky, N., and Meister, G. (2008). A Human snoRNA with MicroRNA-Like Functions. Molecular Cell **32,** 519-528.

Espinosa, A., Zhou, W., Ek, M., Hedlund, M., Brauner, S., Popovic, K., Horvath, L., Wallerskog, T., Oukka, M., Nyberg, F., Kuchroo, V.K., and Wahren-Herlenius, M. (2006). The Sjögren's Syndrome-Associated Autoantigen Ro52

Is an E3 Ligase That Regulates Proliferation and Cell Death. The Journal of Immunology **176,** 6277-6285.

Fabini, G., Rutjes, S.A., Zimmermann, C., Pruijn, G.J.M., and Steiner, G. (2000). Analysis of the molecular composition of Ro ribonucleoprotein complexes. Identification of novel Y RNA-binding proteins. European Journal of Biochemistry **267,** 2778-2789.

Fabini, G., Raijmakers, R., Hayer, S., Fouraux, M.A., Pruijn, G.J., and Steiner, G. (2001). The heterogeneous nuclear ribonucleoproteins I and K interact with a subset of the ro ribonucleoprotein-associated Y RNAs in vitro and in vivo. The Journal of Biological Chemistry **276,** 20711-20718.

Fais, S., O'Driscoll, L., Borras, F.E., Buzas, E., Camussi, G., Cappello, F., Carvalho, J., Cordeiro Da Silva, A., Del Portillo, H., El Andaloussi, S., Ficko Trček, T., Furlan, R., Hendrix, A., Gursel, I., Kralj-Iglic, V., Kaeffer, B., Kosanovic, M., Lekka, M.E., Lipps, G., Logozzi, M., Marcilla, A., Sammar, M., Llorente, A., Nazarenko, I., Oliveira, C., Pocsfalvi, G., Rajendran, L., Raposo, G., Rohde, E., Siljander, P., Van Niel, G., Vasconcelos, M.H., Yáñez-Mó, M., Yliperttula, M.L., Zarovni, N., Zavec, A.B., and Giebel, B. (2016). Evidence-Based Clinical Use of Nanoscale Extracellular Vesicles in Nanomedicine. ACS nano **10,** 3886-3899.

Falaleeva, M., Pages, A., Matuszek, Z., Hidmi, S., Agranat-Tamir, L., Korotkov, K., Nevo, Y., Eyras, E., Sperling, R., and Stamm, S. (2016). Dual function of C/D box small nucleolar RNAs in rRNA modification and alternative pre-mRNA splicing. Proceedings of the National Academy of Sciences of the United States of America **113,** E1625-E1634.

Farris, A.D., Gross, J.K., Hanas, J.S., and Harley, J.B. (1996). Genes for murine Y1 and Y3 Ro RNAs have class 3 RNA polymerase III promoter structures and are unlinked on mouse chromosome 6. Gene **174,** 35-42.

Farris, A.D., Puvion-Dutilleul, F., Puvion, E., Harley, J.B., and Lee, L.A. (1997). The ultrastructural localization of 60-kDa Ro protein and human cytoplasmic RNAs: Association with novel electron-dense bodies. Proceedings of the National Academy of Sciences of the United States of America **94,** 3040-3045.

Farris, A.D., Koelsch, G., Pruijn, G.J., van Venrooij, W.J., and Harley, J.B. (1999). Conserved features of Y RNAs revealed by automated phylogenetic secondary structure analysis. Nucleic Acids Research **27,** 1070-1078.

Filipowicz, W., Bhattacharyya, S.N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: Are the answers in sight? Nature Reviews Genetics **9,** 102-114.

Fiorentino, D.F., Presby, M., Baer, A.N., Petri, M., Rieger, K.E., Soloski, M., Rosen, A., Mammen, A.L., Christopher-Stine, L., and Casciola-Rosen, L. (2016). PUF60: a prominent new target of the autoimmune response in dermatomyositis and Sjögren's syndrome. Annals of the Rheumatic Diseases **75,** 1145-1151.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature **391,** 806-811.

Fouraux, M.A., Bouvet, P., Verkaart, S., van Venrooij, W.J., and Pruijn, G.J. (2002). Nucleolin associates with a subset of the human Ro ribonucleoprotein complexes. Journal of Molecular Biology **320,** 475-488.

Fragkos, M., Ganier, O., Coulombe, P., and Méchali, M. (2015). DNA replication origin activation in space and time. Nature Reviews Molecular Cell Biology **16,** 360.

Frank, D.N., and Pace, N.R. (1998). Ribonuclease P: Unity and Diversity in a tRNA Processing Ribozyme. Annual Review of Biochemistry **67,** 153-180.

Frank, M.B., McCubbin, V.R., and Heldermon, C. (1995). Expression and DNA binding of the human 52 kDa Ro/SSA autoantigen. The Biochemical Journal **305 (Pt 2),** 359-362.

Fu, H., Feng, J., Liu, Q., Sun, F., Tie, Y., Zhu, J., Xing, R., Sun, Z., and Zheng, X. (2009). Stress induces tRNA cleavage by angiogenin in mammalian cells. FEBS Letters **583,** 437-442.

Fuchs, G., Stein, A.J., Fu, C., Reinisch, K.M., and Wolin, S.L. (2006). Structural and biochemical basis for misfolded RNA recognition by the Ro autoantigen. Nature Structural & Molecular Biology **13,** 1002-1009.

Gallois-Montbrun, S., Holmes, R.K., Swanson, C.M., Fernández-Ocaña, M., Byers, H.L., Ward, M.A., and Malim, M.H. (2008). Comparison of Cellular Ribonucleoprotein Complexes Associated with the APOBEC3F and APOBEC3G Antiviral Proteins. Journal of Virology **82,** 5636-5642.

Garcia, E.L., Onafuwa-Nuga, A., Sim, S., King, S.R., Wolin, S.L., and Telesnitsky, A. (2009). Packaging of host mY RNAs by murine leukemia virus may occur early in Y RNA biogenesis. Journal of Virology **83,** 12526-12534.

Gardiner, T.J., Christov, C.P., Langley, A.R., and Krude, T. (2009). A conserved motif of vertebrate Y RNAs essential for chromosomal DNA replication. RNA **15,** 1375-1385.

Gendron, M., Roberge, D., and Boire, G. (2001). Heterogeneity of human Ro ribonucleoproteins (RNPS): nuclear retention of Ro RNPS containing the human hY5 RNA in human and mouse cells. Clinical and Experimental Immunology **125,** 162-168.

Gerstberger, S., Hafner, M., and Tuschl, T. (2014). A census of human RNA-binding proteins. Nature Reviews Genetics **15,** 829-845.

Ghildiyal, M., Xu, J., Seitz, H., Weng, Z., and Zamore, P.D. (2010). Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway. RNA **16,** 43-56.

Ginisty, H., Sicard, H., Roger, B., and Bouvet, P. (1999). Structure and functions of nucleolin. Journal of cell science **112,** 761-772.

Glisovic, T., Bachorik, J.L., Yong, J., and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. FEBS letters **582,** 1977-1986.

Gottlieb, E., and Steitz, J.A. (1989). Function of the mammalian La protein: evidence for its action in transcription termination by RNA polymerase III. The EMBO Journal **8,** 851-861.

Gould, S.J., and Raposo, G. (2013). As we wait: Coping with an imperfect nomenclature for extracellular vesicles. Journal of extracellular vesicles **2**.

Green, C.D., Long, K.S., Shi, H., and Wolin, S.L. (1998). Binding of the 60-kDA Ro autoantigen to Y RNAs: Evidence for recognition in the major groove of a conserved helix. RNA **4,** 750-765.

Griffiths-Jones, S. (2006). miRBase: the microRNA sequence database. Methods in molecular biology **342,** 129-138.

Gruber, A.R., Lorenz, R., Bernhart, S.H., Neuböck, R., and Hofacker, I.L. (2008). The Vienna RNA websuite. Nucleic acids research **36,** W70-74.

Ha, M., and Kim, V.N. (2014). Regulation of microRNA biogenesis. Nature Reviews Molecular Cell Biology **15,** 509-524.

Hafner, M., Renwick, N., Brown, M., Mihailovic, A., Holoch, D., Lin, C., Pena, J.T., Nusbaum, J.D., Morozov, P., Ludwig, J., Ojo, T., Luo, S., Schroth, G., and Tuschl, T. (2011). RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. RNA **17,** 1697-1712.

Haiser, H.J., Karginov, F.V., Hannon, G.J., and Elliot, M.A. (2008). Developmentally regulated cleavage of tRNAs in the bacterium Streptomyces coelicolor. Nucleic Acids Research **36,** 732-741.

Hajdin, C.E., Bellaousov, S., Huggins, W., Leonard, C.W., Mathews, D.H., and Weeks, K.M. (2013). Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. Proceedings of the National Academy of Sciences **110,** 5498-5503.

Hall, A.E. (2013). Biogenesis of Y RNA derived small RNAs.

Hall, A.E., and Dalmay, T. (2013). Discovery of novel small RNAs in the quest to unravel genome complexity. Biochemical Society transactions **41,** 866-870.

Hall, A.E., Turnbull, C., and Dalmay, T. (2013). Y RNAs: Recent developments. Biomolecular concepts **4,** 103-110.

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science (New York, N.Y.) **286,** 950-952.

Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. Nature **404,** 293-296.

Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H., and Kim, V.N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. Genes and Development **18,** 3016-3027.

Han, Y., Donovan, J., Rath, S., Whitney, G., Chitrakar, A., and Korennykh, A. (2014). Structure of Human RNase L Reveals the Basis for Regulated RNA Decay in the IFN Response. Science **343,** 1244-1248.

Haussecker, D., Huang, Y., Lau, A., Parameswaran, P., Fire, A.Z., and Kay, M.A. (2010). Human tRNA-derived small RNAs in the global regulation of RNA silencing. RNA **16,** 673-695.

Heichman, K., and Roberts, J. (1994). Rules to replicate by. Cell 79, 557–562.

Hendrick, J.P., Wolin, S.L., Rinke, J., Lerner, M.R., and Steitz, J.A. (1981). Ro small cytoplasmic ribonucleoproteins are a subclass of La ribonucleoproteins: Further characterization of the Ro and La small ribonucleoproteins from uninfected mammalian cells. Molecular and Cellular Biology **1,** 1138-1149.

Hizir, Z., Bottini, S., Grandjean, V., Trabucchi, M., and Repetto, E. (2017). RNY (Y RNA)-derived small RNAs regulate cell death and inflammation in monocytes/macrophages. Cell Death and Disease **8**.

Hoagland, M.B., Stephenson, M.L., Scott, J.F., Hecht, L.I., and Zamecnik, P.C. (1958). A soluble ribonucleic acid intermediate in protein synthesis. The Journal of Biological Chemistry **231,** 241-257.

Hofacker, I.L. (2003). Vienna RNA secondary structure server. Nucleic acids research **31,** 3429-3431.

Hogg, J.R., and Collins, K. (2007). Human Y5 RNA specializes a Ro ribonucleoprotein for 5S ribosomal RNA quality control. Genes & Development **21,** 3067-3072.

Holcik, M., and Sonenberg, N. (2005). Translational control in stress and apoptosis. Nature Reviews Molecular Cell Biology **6,** 318-327.

Horos, R., Alleaume, A.-M., Kleinendorst, R., Tarafder, A.K., Schwarzl, T., Zielonka, E.M., Adak, A., Castello, A., Huber, W., Sachse, C., and Hentze, M.W. (2017). The small non-coding vault RNA1-1 acts as a riboregulator of autophagy. bioRxiv.

Houbaviy, H.B., Murray, M.F., and Sharp, P.A. (2003). Embryonic Stem Cell-Specific MicroRNAs. Developmental Cell **5,** 351-358.

Hovanessian, A.G., Brown, R.E., and Kerr, I.M. (1977). Synthesis of low molecular weight inhibitor of protein synthesis with enzyme from interferon-treated cells. Nature **268,** 537.

Hung, T., Pratt, G.A., Sundararaman, B., Townsend, M.J., Chaivorapol, C., Bhangale, T., Graham, R.R., Ortmann, W., Criswell, L.A., Yeo, G.W., and Behrens, T.W. (2015). The Ro60 autoantigen binds endogenous retroelements and regulates inflammatory gene expression. Science **350,** 455-459.

Huntzinger, E., and Izaurralde, E. (2011). Gene silencing by microRNAs: Contributions of translational repression and mRNA decay. Nature Reviews Genetics **12,** 99-110.

Hussain, S., Sajini, Abdulrahim A., Blanco, S., Dietmann, S., Lombard, P., Sugimoto, Y., Paramor, M., Gleeson, Joseph G., Odom, Duncan T., Ule, J., and Frye, M. (2013). NSun2-Mediated Cytosine-5 Methylation of Vault Noncoding RNA Determines Its Processing into Regulatory Small RNAs. Cell Reports **4,** 255-261.

Hutvagner, G., and Simard, M.J. (2008). Argonaute proteins: Key players in RNA silencing. Nature Reviews Molecular Cell Biology **9,** 22-32.

Hutvágner, G., McLachlan, J., Pasquinelli, A.E., Bálint, É., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme dicer in the maturation of the let-7 small temporal RNA. Science **293,** 834-838.

International Human Genome Sequencing Consortium. (2004). Finishing the euchromatic sequence of the human genome. Nature **431,** 931-945.

Ishikawa, T., Haino, A., Seki, M., Terada, H., and Nashimoto, M. (2017). The Y4-RNA fragment, a potential diagnostic marker, exists in saliva. Non-coding RNA Research **2,** 122-128.

Ivanov, P., Emara, Mohamed M., Villen, J., Gygi, Steven P., and Anderson, P. (2011). Angiogenin-Induced tRNA Fragments Inhibit Translation Initiation. Molecular Cell **43,** 613-623.

Jayaprakash, A.D., Jabado, O., Brown, B.D., and Sachidanandam, R. (2011). Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. Nucleic acids research **39**.

Jöchl, C., Rederstorff, M., Hertel, J., Stadler, P.F., Hofacker, I.L., Schrettl, M., Haas, H., and Hüttenhofer, A. (2008). Small ncRNA transcriptome analysis from Aspergillus fumigatus suggests a novel mechanism for regulation of protein synthesis. Nucleic Acids Research **36,** 2677-2689.

Johnstone, R.M., Adam, M., Hammond, J.R., Orr, L., and Turbide, C. (1987). Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes). The Journal of Biological Chemistry **262,** 9412-9420.

Jorgensen, R. (1990). Altered gene expression in plants due totrans interactions between homologous genes. Trends in Biotechnology **8,** 340-344.

Katibah, G.E., Qin, Y., Sidote, D.J., Yao, J., Lambowitz, A.M., and Collins, K. (2014). Broad and adaptable RNA structure recognition by the human interferon-induced tetratricopeptide repeat protein IFIT5. Proceedings of the National Academy of Sciences of the United States of America **111,** 12025-12030.

Kato, N., Hoshino, H., and Harada, F. (1982). Nucleotide sequence of 4.5S RNA (C8 or hY5) from HeLa cells. Biochemical and Biophysical Research communications **108,** 363-370.

Kawashima, C.G., Yoshimoto, N., Maruyama-Nakashita, A., Tsuchiya, Y.N., Saito, K., Takahashi, H., and Dalmay, T. (2009). Sulphur starvation induces the expression of microRNA-395 and one of its target genes but in different cell types. Plant Journal **57,** 313-321.

Keam, S.P., and Hutvagner, G. (2015). tRNA-Derived Fragments (tRFs): Emerging New Roles for an Ancient RNA in the Regulation of Gene Expression. Life **5,** 1638-1651.

Kedersha, N.L., and Rome, L.H. (1986). Isolation and characterization of a novel ribonucleoprotein particle: Large structures contain a single species of small RNA. Journal of Cell Biology **103,** 699-709.

Kelekar, A., Saitta, M.R., and Keene, J.D. (1994). Molecular composition of Ro small ribonucleoprotein complexes in human cells. Intracellular localization of the 60- and 52-kD proteins. The Journal of Clinical Investigation **93,** 1637-1644.

Ketting, R.F., Fischer, S.E.J., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H.A. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes and Development **15,** 2654-2659.

Kheir, E.G.A., and Krude, T. (2017). Non-coding Y RNAs associate with early replicating euchromatin concordantly with the origin recognition complex (ORC). Journal of Cell Science.

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell **115,** 209-216.

Kickhoefer, V.A., Rajavel, K.S., Scheffer, G.L., Dalton, W.S., Scheper, R.J., and Rome, L.H. (1998). Vaults are up-regulated in multidrug-resistant cancer cell lines. Journal of Biological Chemistry **273,** 8971-8974.

Kim, E., Lee, S., Mian, M.F., Yun, S.U., Song, M., Yi, K.-S., Ryu, S.H., and Suh, P.-G. (2006). Crosstalk between Src and major vault protein in epidermal growth factor-dependent cell signalling. The FEBS Journal **273,** 793-804.

Kim Kyoung, M., Abdelmohsen, K., Mustapic, M., Kapogiannis, D., and Gorospe, M. (2017). RNA in extracellular vesicles. Wiley Interdisciplinary Reviews: RNA **8,** e1413.

Kim, V.N., Han, J., and Siomi, M.C. (2009). Biogenesis of small RNAs in animals. Nature Reviews Molecular Cellular Biology **10,** 126-139.

Kinoshita, T., Yip, K.W., Spence, T., and Liu, F.-F. (2016). MicroRNAs in extracellular vesicles: potential cancer biomarkers. Journal Of Human Genetics **62,** 67.

Kishore, S., and Stamm, S. (2006). The snoRNA HBII-52 Regulates Alternative Splicing of the Serotonin Receptor 2C. Science **311,** 230-232.

Kishore, S., Khanna, A., Zhang, Z., Hui, J., Balwierz, P.J., Stefan, M., Beach, C., Nicholls, R.D., Zavolan, M., and Stamm, S. (2010). The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing. Human Molecular Genetics **19,** 1153-1164.

Kiss, T. (2001). Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs. The EMBO Journal **20,** 3617-3622.

Kiss, T. (2002). Small Nucleolar RNAs: An Abundant Group of Noncoding RNAs with Diverse Cellular Functions. Cell **109,** 145-148.

Kobayashi, S., Hoshino, T., Hiwasa, T., Satoh, M., Rahmutulla, B., Tsuchida, S., Komukai, Y., Tanaka, T., Matsubara, H., Shimada, H., Nomura, F., and Matsushita, K. (2016). Anti-FIRs (PUF60) auto-antibodies are detected in the sera of early-stage colon cancer patients. Oncotarget **7,** 82493-82503.

Köhn, M., Pazaitis, N., and Hüttelmaier, S. (2013). Why Y RNAs? About Versatile RNAs and Their Functions. Biomolecules **3,** 143.

Köhn, M., Lederer, M., Wächter, K., and Hüttelmaier, S. (2010). Near-infrared (NIR) dye-labeled RNAs identify binding of ZBP1 to the noncoding Y3-RNA. RNA **16,** 1420-1428.

Köhn, M., Ihling, C., Sinz, A., Krohn, K., and Hüttelmaier, S. (2015). The Y3** ncRNA promotes the 3' end processing of histone mRNAs. Genes & Development **29,** 1998-2003.

Kolb, S.J., Battle, D.J., and Dreyfuss, G. (2007). Molecular functions of the SMN complex. Journal of Child Neurology **22,** 990-994.

Kowalski, M.P., and Krude, T. (2015). Functional roles of non-coding Y RNAs. The International Journal of Biochemistry & Cell Biology **66,** 20-29.

Kowalski, M.P., Baylis, H.A., and Krude, T. (2015). Non-coding stem-bulge RNAs are required for cell proliferation and embryonic development in C. elegans. Journal of Cell Science **128,** 2118-2129.

Krämer, A. (1996). The Structure and Function of Proteins Involved in Mammalian Pre-mRNA Splicing. Annual Review of Biochemistry **65,** 367-409.

Krude, T., Jackman, M., Pines, J., and Laskey, R.A. (1997). Cyclin/Cdk-dependent initiation of DNA replication in a human cell-free system. Cell **88,** 109-119.

Krude, T. (2000). Initiation of human DNA replication in vitro using nuclei from cellsarrested at an initiation-competent state. Journal of Biological Chemistry **275**, 13699–13707.

Krude, T., Christov, C.P., Hyrien, O., and Marheineke, K. (2009). Y RNA functions at the initiation step of mammalian chromosomal DNA replication. Journal of Cell Science **122,** 2836-2845.

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. Cell **31,** 147-157.

Kültz, D. (2005). Molecular and evolutionary basis of the cellular stress response. In Annual Review of Physiology, pp. 225-257.

Kumar, P., Kuscu, C., and Dutta, A. (2016). Biogenesis and Function of Transfer RNA-Related Fragments (tRFs). Trends in Biochemical Sciences **41,** 679-689.

Kumar, P., Anaya, J., Mudunuri, S.B., and Dutta, A. (2014). Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. BMC Biology **12,** 78.

Kwak, P.B., and Tomari, Y. (2012). The N domain of Argonaute drives duplex unwinding during RISC assembly. Nature Structural & Molecular Biology **19,** 145-151.

Labbé, J.C., Hekimi, S., and Rokeach, L.A. (1999). The levels of the RoRNP-associated Y RNA are dependent upon the presence of ROP-1, the Caenorhabditis elegans Ro60 protein. Genetics **151,** 143-150.

Langley, A.R., Chambers, H., Christov, C.P., and Krude, T. (2010). Ribonucleoprotein particles containing non-coding Y RNAs, Ro60, La and nucleolin are not required for Y RNA function in DNA replication. PloS one **5,** e13673.

Langmead, B. (2010). Aligning short sequencing reads with Bowtie. Current Protocols in Bioinformatics.

Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science **294,** 858-862.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell **75,** 843-854.

Lee, S.R., and Collins, K. (2005). Starvation-induced cleavage of the tRNA anticodon loop in Tetrahymena thermophila. Journal of Biological Chemistry **280,** 42744-42749.

Lee, Y.S., Shibata, Y., Malhotra, A., and Dutta, A. (2009). A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). Genes and Development **23,** 2639-2649.

Lerner, M.R., and Steitz, J. (1979). Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. Proceedings of the National Academy of Sciences of the United States of America **76,** 5495-5499.

Lerner, M.R., Boyle, J.A., Hardin, J.A., and Steitz, J.A. (1981a). Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus. Science **211,** 400-402.

Lerner, M.R., Andrews, N.C., Miller, G., and Steitz, J.A. (1981b). Two small RNAs encoded by Epstein-Barr virus and complexed with protein are precipitated by antibodies from patients with systemic lupus erythematosus. Proceedings of the National Academy of Sciences of the United States of America **78,** 805-809.

Leung, A.K.L., and Sharp, P.A. (2007). microRNAs: A Safeguard against Turmoil? Cell **130,** 581-585.

Leung, A.K.L., and Sharp, P.A. (2010). MicroRNA Functions in Stress Responses. Molecular Cell **40,** 205-215.

Leung, K.N., Vallero, R.O., DuBose, A.J., Resnick, J.L., and LaSalle, J.M. (2009). Imprinting regulates mammalian snoRNA-encoding chromatin decondensation and neuronal nucleolar size. Human Molecular Genetics **18,** 4227-4238.

Li, Y., Banerjee, S., Wang, Y., Goldstein, S.A., Dong, B., Gaughan, C., Silverman, R.H., and Weiss, S.R. (2016). Activation of RNase L is dependent on OAS3 expression during infection with diverse human viruses. Proceedings of the National Academy of Sciences of the United States of America **113,** 2241-2246.

Li, Z., Ender, C., Meister, G., Moore, P.S., Chang, Y., and John, B. (2012). Extensive terminal and asymmetric processing of small RNAs from rRNAs, snoRNAs, snRNAs, and tRNAs. Nucleic Acids Research **40,** 6787-6799.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology **15,** 550.

Lu, Z.H., Books, J.T., and Ley, T.J. (2005). YB-1 is important for late-stage embryonic development, optimal cellular stress responses, and the prevention of premature senescence. Molecular and Cellular Biology **25,** 4625-4637.

Lukong, K.E., Chang, K.-w., Khandjian, E.W., and Richard, S. (2008). RNA-binding proteins in human genetic disease. Trends in Genetics **24,** 416-425.

Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: Modular design for efficient function. Nature Reviews Molecular Cell Biology **8,** 479-490.

Maas, S.L.N., Breakefield, X.O., and Weaver, A.M. (2017). Extracellular Vesicles: Unique Intercellular Delivery Vehicles. Trends in cell biology **27,** 172-188.

Mannoor, K., Liao, J., and Jiang, F. (2012). Small nucleolar RNAs in cancer. Biochimica et biophysica acta **1826,** 10.1016/j.bbcan.2012.1003.1005.

Maraia, R., Sakulich, A.L., Brinkmann, E., and Green, E.D. (1996). Gene encoding human Ro-associated autoantigen Y5 RNA. Nucleic Acids Research **24,** 3552-3559.

Maraia, R.J., and Lamichhane, T.N. (2011). 3′ processing of eukaryotic precursor tRNAs. Wiley Interdisciplinary Reviews: RNA **2,** 362-375.

Maraia, R.J., Kenan, D.J., and Keene, J.D. (1994a). Eukaryotic transcription termination factor La mediates transcript release and facilitates reinitiation by RNA polymerase III. Molecular and Cellular Biology **14,** 2147-2158.

Maraia, R.J., Sasaki-Tozawa, N., Driscoll, C.T., Green, E.D., and Darlington, G.J. (1994b). The human Y4 small cytoplasmic RNA gene is controlled by upstream elements and resides on chromosome 7 with all other hY scRNA genes. Nucleic Acids Research **22,** 3045-3052.

Martens-Uzunova, E.S., Jalava, S.E., Dits, N.F., van Leenders, G.J.L.H., Møller, S., Trapman, J., Bangma, C.H., Litman, T., Visakorpi, T., and Jenster, G. (2011). Diagnostic and prognostic signatures from the small non-coding RNA transcriptome in prostate cancer. Oncogene **31,** 978.

Martinez, F.J., Lee, J.H., Lee, J.E., Blanco, S., Nickerson, E., Gabrie, S., Frye, M., Al-Gazali, L., and Gleeson, J.G. (2012). Whole exome sequencing identifies a splicing mutation in NSUN2 as a cause of a Dubowitz-like syndrome. Journal of Medical Genetics **49,** 380-385.

Matera, A.G., Frey, M.R., Margelot, K., and Wolin, S.L. (1995). A perinucleolar compartment contains several RNA polymerase III transcripts as well as the polypyrimidine tract-binding protein, hnRNP I. The Journal of Cell Biology **129,** 1181-1193.

Mattick, J.S. (2001). Non-coding RNAs: the architects of eukaryotic complexity. EMBO Reports **2,** 986-991.

Mattick, J.S. (2003). Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. Bioessays **25,** 930-939.

Mattick, J.S. (2004). RNA regulation: a new genetics? Nature Reviews Genetics **5,** 316-323.

Maxwell, E., and Fournier, M. (1995). The Small Nucleolar RNAs. Annual Review of Biochemistry **64,** 897-934.

Meiri, E., Levy, A., Benjamin, H., Ben-David, M., Cohen, L., Dov, A., Dromi, N., Elyakim, E., Yerushalmi, N., Zion, O., Lithwick-Yanai, G., and Sitbon, E. (2010). Discovery of microRNAs and other small RNAs in solid tumors. Nucleic acids research **38,** 6234-6246.

Meister, G. (2013). Argonaute proteins: Functional insights and emerging roles. Nature Reviews Genetics **14,** 447-459.

Minciacchi, V.R., Freeman, M.R., and Di Vizio, D. (2015). Extracellular vesicles in cancer: exosomes, microvesicles and the emerging role of large oncosomes. Seminars in cell & developmental biology **40,** 41-51.

Morris, K.V., and Mattick, J.S. (2014). The rise of regulatory RNA. Nature Reviews Genetics **15,** 423.

Mosig, A., Meng, G., Stadler, B.M.R., and Stadler, P.F. (2007). Evolution of the vertebrate Y RNA cluster. Theory in Biosciences **126,** 9-14.

Nandy, C., Mrázek, J., Stoiber, H., Grässer, F.A., Hüttenhofer, A., and Polacek, N. (2009). Epstein–Barr Virus-Induced Expression of a Novel Human Vault RNA. Journal of Molecular Biology **388,** 776-784.

Napoli, C., Lemieux, C., and Jorgensen, R. (1990). Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. The Plant Cell **2,** 279-289.

Nicolas, F.E., Hall, A.E., Csorba, T., Turnbull, C., and Dalmay, T. (2012). Biogenesis of Y RNA-derived small RNAs is independent of the microRNA pathway. FEBS Letters **586,** 1226-1230.

Ninomiya, S., Kawano, M., Abe, T., Ishikawa, T., Takahashi, M., Tamura, M., Takahashi, Y., and Nashimoto, M. (2015). Potential small guide RNAs for tRNase ZL from human plasma, peripheral blood mononuclear cells, and cultured cell lines. PloS one **10,** e0118631.

Nolte-'t Hoen, E.N., Buermans, H.P., Waasdorp, M., Stoorvogel, W., Wauben, M.H., and t Hoen, P.A. (2012). Deep sequencing of RNA from immune cell-derived vesicles uncovers the selective incorporation of small non-coding RNA biotypes with potential regulatory functions. Nucleic acids research **40,** 9272-9285.

Nurse, P. (1994). Ordering S phase and M phase in the cell cycle. Cell 79, 547-550

O'Brien, C.A., and Wolin, S.L. (1994). A possible role for the 60-kD Ro autoantigen in a discard pathway for defective 5S rRNA precursors. Genes and Development **8,** 2891-2903.

O'Brien, C.A., Margelot, K., and Wolin, S.L. (1993). Xenopus Ro ribonucleoproteins: members of an evolutionarily conserved class of cytoplasmic ribonucleoproteins. Proceedings of the National Academy of Sciences of the United States of America **90,** 7250-7254.

Ogata, J., Sugiura, Y., Kanai, A., Tanaka, M., Matsui, H., Ohtsuka, M., Inaba, T., Otsuka, M., and Kotani, A. (2018). Identification of PTBP1 responsible for caspase dependent Y RNA cleavage. bioRxiv.

Okamura, K., Phillips, M.D., Tyler, D.M., Duan, H., Chou, Y.T., and Lai, E.C. (2008). The regulatory activity of microRNA* species has substantial influence on microRNA and 3′ UTR evolution. Nature Structural and Molecular Biology **15,** 354-363.

Ottosson, L., Hennig, J., Espinosa, A., Brauner, S., Wahren-Herlenius, M., and Sunnerhagen, M. (2006). Structural, functional and immunologic characterization of folded subdomains in the Ro52 protein targeted in Sjögren's syndrome. Molecular Immunology **43,** 588-598.

Page-McCaw, P.S., Amonlirdviman, K., and Sharp, P.A. (1999). PUF60: a novel U2AF65-related splicing activity. RNA (New York, N.Y.) **5,** 1548-1560.

Pakos-Zebrucka, K., Koryga, I., Mnich, K., Ljujic, M., Samali, A., and Gorman, A.M. (2016). The integrated stress response. EMBO reports **17,** 1374-1395.

Palade, G.E. (1955). A small particulate component of the cytoplasm. The Journal of Biophysical and Biochemical Cytology **1,** 59-68.

Palazzo, A.F., and Lee, E.S. (2015). Non-coding RNA: what is functional and what is junk? Frontiers in Genetics **6,** 2.

Pall, G.S., and Hamilton, A.J. (2008). Improved northern blot method for enhanced detection of small RNA. Nature Protocols **3,** 1077-1084.

Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Muller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., and Ruvkun, G. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. Nature **408,** 86-89.

Patterson, D.G., Roberts, J.T., King, V.M., Houserova, D., Barnhill, E.C., Crucello, A., Polska, C.J., Brantley, L.W., Kaufman, G.C., Nguyen, M., Santana, M.W., Schiller, I.A., Spicciani, J.S., Zapata, A.K., Miller, M.M., Sherman, T.D., Ma, R., Zhao, H., Arora, R., Coley, A.B., Zeidan, M.M., Tan, M., Xi, Y., and Borchert, G.M. (2017). Human snoRNA-93 is processed into a microRNA-like RNA that promotes breast cancer cell invasion. npj Breast Cancer **3,** 25.

Pederson, T. (2010). Regulatory RNAs derived from transfer RNA? RNA **16,** 1865-1869.

Peek, R., Pruijn, G.J.M., Van der Kemp, A.J.W., and Van Venrooij, W.J. (1993). Subcellular distribution of Ro ribonucleoprotein complexes and their constituents. Journal of Cell Science **106,** 929-935.

Pereira, B., Billaud, M., and Almeida, R. (2017). RNA-Binding Proteins in Cancer: Old Players and New Actors. Trends in cancer **3,** 506-528.

Perreault, J., Perreault, J.P., and Boire, G. (2007). Ro-associated Y RNAs in metazoans: evolution and diversification. Molecular Biology and Evolution **24,** 1678-1689.

Perreault, J., Noël, J.-F., Brière, F., Cousineau, B., Lucier, J.-F., Perreault, J.-P., and Boire, G. (2005). Retropseudogenes derived from the human Ro/SS-A autoantigen-associated hY RNAs. Nucleic Acids Research **33,** 2032-2041.

Persson, H., Kvist, A., Vallon-Christersson, J., Medstrand, P., Borg, Å., and Rovira, C. (2009). The non-coding RNA of the multidrug resistance-linked vault particle encodes multiple regulatory small RNAs. Nature Cell Biology **11,** 1268.

Pfeffer, S., Zavolan, M., Grasser, F.A., Chien, M., Russo, J.J., Ju, J., John, B., Enright, A.J., Marks, D., Sander, C., and Tuschl, T. (2004). Identification of virus-encoded microRNAs. Science **304,** 734-736.

Pospiech, H., Grosse, F., and Pisani, F.M. (2010). The initiation step of eukaryotic DNA replication. Sub-Cellular Biochemistry **50,** 79-104.

Prüfer, K., Stenzel, U., Dannemann, M., Green, R.E., Lachmann, M., and Kelso, J. (2008). PatMaN: Rapid alignment of short sequences to large databases. Bioinformatics **24,** 1530-1531.

Pruijn, G.J., Slobbe, R.L., and van Venrooij, W.J. (1991). Analysis of protein--RNA interactions within Ro ribonucleoprotein complexes. Nucleic Acids Research **19,** 5173-5180.

Pruijn, G.J.M., Wingens, P.A.E.T.M., Peters, S.L.M., Thijssen, J.H., and van Venrooij, W.J. (1993). Ro RNP associated Y RNAs are highly conserved among mammals. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression **1216,** 395-401.

Qi, Y., Tu, J., Cui, L., Guo, X., Shi, Z., Li, S., Shi, W., Shan, Y., Ge, Y., Shan, J., Wang, H., and Lu, Z. (2010). High-Throughput Sequencing of MicroRNAs in Adenovirus Type 3 Infected Human Laryngeal Epithelial Cells. Journal of Biomedicine and Biotechnology **2010,** 915980.

Quinlan, A.R. (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. Current protocols in bioinformatics **47,** 11.12.11-34.

Rao, P.N., and Johnson, R.T. (1970). Mammalian Cell Fusion : Studies on the Regulation of DNA Synthesis and Mitosis. Nature **225,** 159.

Ratajczak, J., Miekus, K., Kucia, M., Zhang, J., Reca, R., Dvorak, P., and Ratajczak, M.Z. (2006). Embryonic stem cell-derived microvesicles reprogram hematopoietic progenitors: evidence for horizontal transfer of mRNA and protein delivery. Leukemia **20,** 847-856.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. Nature **403,** 901-906.

Rinke, J., and Steitz, J.A. (1982). Precursor molecules of both human 5S ribosomal RNA and transfer RNAs are bound by a cellular protein reactive with anti-La Lupus antibodies. Cell **29,** 149-159.

Rivas, F.V., Tolia, N.H., Song, J.J., Aragon, J.P., Liu, J., Hannon, G.J., and Joshua-Tor, L. (2005). Purified Argonaute2 and an siRNA form recombinant human RISC. Nature structural & molecular biology **12,** 340-349.

Rosa, M.D., Gottlieb, E., Lerner, M.R., and Steitz, J.A. (1981). Striking similarities are exhibited by two small Epstein-Barr virus-encoded ribonucleic acids and the adenovirus-associated ribonucleic acids VAI and VAII. Molecular and Cellular Biology **1,** 785-796.

Rother, S., and Meister, G. (2011). Small RNAs derived from longer non-coding RNAs. Biochimie **93,** 1905-1915.

Rutjes, S.A., van der Heijden, A., Utz, P.J., van Venrooij, W.J., and Pruijn, G.J. (1999a). Rapid nucleolytic degradation of the small cytoplasmic Y RNAs during apoptosis. Journal of Biological Chemistry **274,** 24799-24807.

Rutjes, S.A., Utz, P.J., van der Heijden, A., Broekhuis, C., van Venrooij, W.J., and Pruijn, G.J.M. (1999b). The La (SS-B) autoantigen, a key protein in RNA biogenesis, is dephosphorylated and cleaved early during apoptosis. Cell death and differentiation **6,** 976.

Schiffer, S., Rösch, S., and Marchfelder, A. (2002). Assigning a function to a conserved group of proteins: the tRNA 3′-processing enzymes. The EMBO Journal **21,** 2769-2777.

Schotte, D., Chau, J.C., Sylvester, G., Liu, G., Chen, C., van der Velden, V.H., Broekhuis, M.J., Peters, T.C., Pieters, R., and den Boer, M.L. (2009). Identification of new microRNA genes and aberrant microRNA profiles in childhood acute lymphoblastic leukemia. Leukemia **23,** 313-322.

Schubert, T., Pusch, Miriam C., Diermeier, S., Benes, V., Kremmer, E., Imhof, A., and Längst, G. (2012). Df31 Protein and snoRNAs Maintain Accessible Higher-Order Structures of Chromatin. Molecular Cell **48,** 434-444.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell **115,** 199-208.

Shen, Y., Yu, X., Zhu, L., Li, T., Yan, Z., and Guo, J. (2018). Transfer RNA-derived fragments and tRNA halves: biogenesis, biological functions and their roles in diseases. Journal of Molecular Medicine.

Shi, H., O'Brien, C.A., Van Horn, D.J., and Wolin, S.L. (1996). A misfolded form of 5S rRNA is complexed with the Ro and La autoantigens. RNA **2,** 769-784.

Shimamoto, Y., Sumizawa, T., Haraguchi, M., Gotanda, T., Jueng, H.C., Furukawa, T., Sakata, R., and Akiyama, S.I. (2006). Direct activation of the human major vault protein gene by DNA-damaging agents. Oncology Reports **15,** 645-652.

Shukla, S., and Parker, R. (2017). PARN modulates Y RNA stability and its 3'-end formation. Molecular and Cellular Biology **37**.

Silverman, R.H. (2007). Viral Encounters with 2′,5′-Oligoadenylate Synthetase and RNase L during the Interferon Antiviral Response. Journal of Virology **81,** 12720-12729.

Sim, S., and Wolin, S.L. (2011). Emerging roles for the Ro 60-kDa autoantigen in noncoding RNA metabolism. Wiley interdisciplinary reviews. RNA **2,** 686-699.

Sim, S., Weinberg, D.E., Fuchs, G., Choi, K., Chung, J., and Wolin, S.L. (2009). The Subcellular Distribution of an RNA Quality Control Protein, the Ro Autoantigen, Is Regulated by Noncoding Y RNA Binding. Molecular Biology of the Cell **20,** 1555-1564.

Sim, S., Yao, J., Weinberg, D.E., Niessen, S., Yates, J.R., 3rd, and Wolin, S.L. (2012). The zipcode-binding protein ZBP1 influences the subcellular location of the Ro 60-kDa autoantigen and the noncoding Y3 RNA. RNA **18,** 100-110.

Simons, F.H., Pruijn, G.J., and van Venrooij, W.J. (1994). Analysis of the intracellular localization and assembly of Ro ribonucleoprotein particles by microinjection into Xenopus laevis oocytes. The Journal of Cell Biology **125,** 981-988.

Simons, F.H., Rutjes, S.A., van Venrooij, W.J., and Pruijn, G.J. (1996). The interactions with Ro60 and La differentially affect nuclear export of hY1 RNA. RNA **2,** 264-273.

Slobbe, R.L., Pluk, W., van Venrooij, W.J., and Pruijn, G.J. (1992). Ro ribonucleoprotein assembly in vitro. Identification of RNA-protein and protein-protein interactions. Journal of Molecular Biology **227,** 361-366.

Slobbe, R.L., Pruijn, G.J., Damen, W.G., van der Kemp, J.W., and van Venrooij, W.J. (1991). Detection and occurrence of the 60- and 52-kD Ro (SS-A) antigens and of autoantibodies against these proteins. Clinical and Experimental Immunology **86,** 99-105.

Smith, J.A., Schmechel, S.C., Williams, B.R.G., Silverman, R.H., and Schiff, L.A. (2005). Involvement of the interferon-regulated antiviral proteins PKR and RNase L in reovirus-induced shutoff of cellular translation. Journal of Virology **79,** 2240-2250.

Sobala, A., and Hutvagner, G. (2013). Small RNAs derived from the 5′ end of tRNA can inhibit protein translation in human cells. RNA Biology **10,** 553-563.

Sorefan, K., Pais, H., Hall, A.E., Kozomara, A., Griffiths-Jones, S., Moulton, V., and Dalmay, T. (2012). Reducing ligation bias of small RNAs in libraries for next generation sequencing. Silence **3,** 4.

Spriggs, K.A., Bushell, M., and Willis, A.E. (2010). Translational Regulation of Gene Expression during Conditions of Cell Stress. Molecular Cell **40,** 228-237.

Srivastava, M., and Pollard, H.B. (1999). Molecular dissection of nucleolin's role in growth and cell proliferation: New insights. FASEB Journal **13,** 1911-1922.

Stadler, P.F., Chen, J.J.L., Hackermüller, J., Hoffmann, S., Horn, F., Khaitovich, P., Kretzschmar, A.K., Mosig, A., Prohaska, S.J., Qi, X., Schutt, K., and Ullmann, K. (2009). Evolution of Vault RNAs. Molecular Biology and Evolution **26,** 1975-1991.

Stefano, J.E. (1984). Purified lupus antigen la recognizes an oligouridylate stretch common to the 3′ termini of RNA polymerase III transcripts. Cell **36,** 145-154.

Stein, A.J., Fuchs, G., Fu, C., Wolin, S.L., and Reinisch, K.M. (2005). Structural insights into RNA quality control: the Ro autoantigen binds misfolded RNAs via its central cavity. Cell **121,** 529-539.

Steiner, E., Holzmann, K., Elbling, L., Micksche, M., and Berger, W. (2006). Cellular functions of vaults and their involvement in multidrug resistance. Current Drug Targets **7,** 923-934.

Steitz, J.A., Berg, C., Hendrick, J.P., La Branche-Chabot, H., Metspalu, A., Rinke, J., and Yario, T. (1988). A 5S rRNA/L5 complex is a precursor to ribosome assembly in mammalian cells. The Journal of Cell Biology **106,** 545-556.

Stocks, M.B., Moxon, S., Mapleson, D., Woolfenden, H.C., Mohorianu, I., Folkes, L., Schwach, F., Dalmay, T., and Moulton, V. (2012). The UEA sRNA workbench: A suite of tools for analysing and visualizing next generation sequencing microRNA and small RNA datasets. Bioinformatics **28,** 2059-2061.

Storz, G. (2002). An expanding universe of noncoding RNAs. Science **296,** 1260-1263.

Stuart, B.D., Choi, J., Zaidi, S., Xing, C., Holohan, B., Chen, R., Choi, M., Dharwadkar, P., Torres, F., Girod, C.E., Weissler, J., Fitzgerald, J., Kershaw, C., Klesney-Tait, J., Mageto, Y., Shay, J.W., Ji, W., Bilguvar, K., Mane, S., Lifton, R.P., and Garcia, C.K. (2015). Exome sequencing links mutations in PARN and RTEL1 with familial pulmonary fibrosis and telomere shortening. Nature Genetics **47,** 512-517.

Sunkar, R., Chinnusamy, V., Zhu, J., and Zhu, J.-K. (2007). Small RNAs as big players in plant abiotic stress responses and nutrient deprivation. Trends in Plant Science **12,** 301-309.

Taft, R.J., Glazov, E.A., Lassmann, T., Hayashizaki, Y., Carninci, P., and Mattick, J.S. (2009). Small RNAs derived from snoRNAs. RNA **15,** 1233-1240.

Takeda, D.Y., and Dutta, A. (2005). DNA replication and progression through S phase. Oncogene **24,** 2827.

Takisawa, H., Mimura, S., and Kubota, Y. (2000). Eukaryotic DNA replication: From pre-replication complex to initiation complex. Current Opinion in Cell Biology **12,** 690-696.

Tanaka, N., Chakravarty, A.K., Maughan, B., and Shuman, S. (2011). Novel Mechanism of RNA Repair by RtcB via Sequential 2′,3′-Cyclic Phosphodiesterase and 3′-Phosphate/5′-Hydroxyl Ligation Reactions. Journal of Biological Chemistry **286,** 43134-43143.

Telonis, A.G., Loher, P., Honda, S., Jing, Y., Palazzo, J., Kirino, Y., and Rigoutsos, I. (2015). Dissecting tRNA-derived fragment complexities using personalized transcriptomes reveals novel fragment classes and unexpected dependencies. Oncotarget **6,** 24797-24822.

Teunissen, S.W., Kruithof, M.J., Farris, A.D., Harley, J.B., Venrooij, W.J., and Pruijn, G.J. (2000). Conserved features of Y RNAs: a comparison of experimentally derived secondary structures. Nucleic Acids Research **28,** 610-619.

Thompson, D.M., and Parker, R. (2009). The RNase Rny1p cleaves tRNAs and promotes cell death during oxidative stress in Saccharomyces cerevisiae. Journal of Cell Biology **185,** 43-50.

Thompson, D.M., Lu, C., Green, P.J., and Parker, R. (2008). tRNA cleavage is a conserved response to oxidative stress in eukaryotes. RNA **14,** 2095-2103.

Tkach, M., and Théry, C. (2016). Communication by Extracellular Vesicles: Where We Are and Where We Need to Go. Cell **164,** 1226-1232.

Tolkach, Y., Stahl, A.F., Niehoff, E.-M., Zhao, C., Kristiansen, G., Müller, S.C., and Ellinger, J. (2017). Y RNA expression predicts survival in bladder cancer patients. BMC Cancer **17,** 749.

Tolkach, Y., Niehoff, E.-M., Stahl, A.F., Zhao, C., Kristiansen, G., Müller, S.C., and Ellinger, J. (2018). Y RNA expression in prostate cancer patients: diagnostic and prognostic implications. World Journal of Urology **36,** 1073-1078.

Turnbull, C. (2014). Sequence and structure requirements of Y RNA-derived small RNA biogenesis (University of East Anglia).

Turner, M., and Díaz-Muñoz, M.D. (2018). RNA-binding proteins control gene expression and cell fate in the immune system Nature Immunology **19,** 120-129.

Valadi, H., Ekstrom, K., Bossios, A., Sjostrand, M., Lee, J.J., and Lotvall, J.O. (2007). Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nature Cell Biology **9,** 654-659.

van Gelder, C.W., Thijssen, J.P., Klaassen, E.C., Sturchler, C., Krol, A., van Venrooij, W.J., and Pruijn, G.J. (1994). Common structural features of the Ro RNP associated hY1 and hY5 RNAs. Nucleic Acids Research **22,** 2498-2506.

Van Horn, D.J., Eisenberg, D., O'Brien, C.A., and Wolin, S.L. (1995). Caenorhabditis elegans embryos contain only one major species of Ro RNP. RNA **1,** 293-303.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J.*, et al.* (2001). The sequence of the human genome. Science **291,** 1304-1351.

Victoria Martinez, B., Dhahbi, J.M., Nunez Lopez, Y.O., Lamperska, K., Golusinski, P., Luczewski, L., Kolenda, T., Atamna, H., Spindler, S.R., Golusinski, W., and Masternak, M.M. (2015). Circulating small non-coding RNA signature in head and neck squamous cell carcinoma. Oncotarget **6,** 19246-19263.

Vihervaara, A., Duarte, F.M., and Lis, J.T. (2018). Molecular mechanisms driving transcriptional stress responses. Nature Reviews Genetics **19,** 385-397.

Vojtech, L., Woo, S., Hughes, S., Levy, C., Ballweber, L., Sauteraud, R.P., Strobl, J., Westerberg, K., Gottardo, R., Tewari, M., and Hladik, F. (2014). Exosomes in human semen carry a distinctive repertoire of small non-coding RNAs with potential regulatory functions. Nucleic Acids Research **42,** 7290-7304.

Vollmar, F., Hacker, C., Zahedi, R.-P., Sickmann, A., Ewald, A., Scheer, U., and Dabauvalle, M.-C. (2009). Assembly of nuclear pore complexes mediated by major vault protein. Journal of Cell Science **122,** 780-786.

Wang, I., Kowalski, M.P., Langley, A.R., Rodriguez, R., Balasubramanian, S., Hsu, S.-T.D., and Krude, T. (2014). Nucleotide Contributions to the Structural Integrity and DNA Replication Initiation Activity of Noncoding Y RNA. Biochemistry **53,** 5848-5863.

Wang, W., Chen, X., Wolin, S.L., and Xiong, Y. (2018). Structural Basis for tRNA Mimicry by a Bacterial Y RNA. Structure **26,** 1635-1644.e1633.

Wassarman, K.M., Zhang, A., and Storz, G. (1999). Small RNAs in Escherichia coli. Trends in Microbiology **7,** 37-45.

Weinstein, L.B., and Steitz, J.A. (1999). Guided tours: from precursor snoRNA to functional snoRNP. Current Opinion in Cell Biology **11,** 378-384.

Wilkinson, K.A., Merino, E.J., and Weeks, K.M. (2006). Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nature Protocols **1,** 1610.

Wittmann, J., and Jäck, H.M. (2010). New surprises from the deep - The family of small regulatory RNAs increases. The Scientific World Journal **10,** 1239-1243.

Wolin, S.L., and Steitz, J.A. (1983). Genes for two small cytoplasmic Ro RNAs are adjacent and appear to be single-copy in the human genome. Cell **32,** 735-744.

Wolin, S.L., and Steitz, J.A. (1984). The Ro small cytoplasmic ribonucleoproteins: Identification of the antigenic protein and its binding site on the Ro RNAs. Proceedings of the National Academy of Sciences of the United States of America **81,** 1996-2000.

Wolin, S.L., and Cedervall, T. (2002). The La protein. In Annual Review of Biochemistry, pp. 375-403.

Wolin, S.L., Belair, C., Boccitto, M., Chen, X., Sim, S., Taylor, D.W., and Wang, H.W. (2013). Non-coding Y RNAs as tethers and gates: Insights from bacteria. RNA Biology **10,** 1602-1608.

Wreschner, D.H., McCauley, J.W., Skehel, J.J., and Kerr, I.M. (1981). Interferon action--sequence specificity of the ppp(A2'p)nA-dependent ribonuclease. Nature **289,** 414-417.

Wurtmann, E.J., and Wolin, S.L. (2010). A role for a bacterial ortholog of the Ro autoantigen in starvation-induced rRNA degradation. Proceedings of the National Academy of Sciences of the United States of America **107,** 4022-4027.

Xiao, Q., Sharp, T.V., Jeffrey, I.W., James, M.C., Pruijn, G.J., van Venrooij, W.J., and Clemens, M.J. (1994). The La antigen inhibits the activation of the interferon-inducible protein kinase PKR by sequestering and unwinding double-stranded RNA. Nucleic Acids Research **22,** 2512-2518.

Xu, P., Billmeier, M., Mohorianu, I., Green, D., Fraser, W.D., and Dalmay, T. (2015). An improved protocol for small RNA library construction using High Definition adapters. Methods in Next Generation Sequencing **2**.

Xue, D., Shi, H., Smith, J.D., Chen, X., Noe, D.A., Cedervall, T., Yang, D.D., Eynon, E., Brash, D.E., Kashgarian, M., Flavell, R.A., and Wolin, S.L. (2003). A lupus-like syndrome develops in mice lacking the Ro 60-kDa protein, a major lupus autoantigen. Proceedings of the National Academy of Sciences of the United States of America **100,** 7503-7508.

Yamakawa, N., Okuyama, K., Ogata, J., Kanai, A., Helwak, A., Takamatsu, M., Imadome, K.-i., Takakura, K., Chanda, B., Kurosaki, N., Yamamoto, H., Ando, K., Matsui, H., Inaba, T., and Kotani, A. (2014). Novel functional small RNAs are selectively loaded onto mammalian Ago1. Nucleic Acids Research **42,** 5289-5301.

Yamasaki, S., Ivanov, P., Hu, G.-f., and Anderson, P. (2009). Angiogenin cleaves tRNA and promotes stress-induced translational repression. The Journal of Cell Biology **185,** 35-42.

Yanez-Mo, M., Siljander, P.R., Andreu, Z., Zavec, A.B., Borras, F.E., Buzas, E.I., Buzas, K., Casal, E., Cappello, F., Carvalho, J., Colas, E., Cordeiro-da Silva, A., Fais, S., Falcon-Perez, J.M., Ghobrial, I.M., Giebel, B., Gimona, M., Graner, M., Gursel, I., Gursel, M., Heegaard, N.H., Hendrix, A., Kierulf, P., Kokubun, K., Kosanovic, M., Kralj-Iglic, V., Kramer-Albers, E.M., Laitinen, S., Lasser, C., Lener, T., Ligeti, E., Line, A., Lipps, G., Llorente, A., Lotvall, J., Mancek-Keber, M., Marcilla, A., Mittelbrunn, M., Nazarenko, I., Nolte-'t Hoen, E.N., Nyman, T.A., O'Driscoll, L., Olivan, M., Oliveira, C., Pallinger, E., Del Portillo, H.A., Reventos, J., Rigau, M., Rohde, E., Sammar, M., Sanchez-Madrid, F., Santarem, N., Schallmoser, K., Ostenfeld, M.S., Stoorvogel, W., Stukelj, R., Van der Grein, S.G., Vasconcelos, M.H., Wauben, M.H., and De Wever, O. (2015). Biological properties of extracellular vesicles and their physiological functions. Journal of Extracellular Vesicles **4,** 27066.

Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell **101,** 25-33.

Zhang, A.T., Langley, A.R., Christov, C.P., Kheir, E., Shafee, T., Gardiner, T.J., and Krude, T. (2011). Dynamic interaction of Y RNAs with chromatin and initiation proteins during human DNA replication. Journal of Cell Science **124,** 2058-2069.

Zhou, A., Paranjape, J., Brown, T.L., Nie, H., Naik, S., Dong, B., Chang, A., Trapp, B., Fairchild, R., Colmenares, C., and Silverman, R.H. (1997). Interferon action and apoptosis are defective in mice devoid of 2′,5′-oligoadenylate-dependent RNase L. The EMBO Journal **16,** 6355-6363.

Zhu, L., Liu, X., Pu, W., and Peng, Y. (2018). tRNA-derived small non-coding RNAs in human disease. Cancer Letters **419,** 1-7.

Appendices

Appendix I

Media, buffers and solutions

Media and solutions for bacteria work

LB -medium	1% (w/v) bacto tryptone (Formedium™) 1% (w/v) NaCl (Fisher Scientific) 0.5% (w/v) yeast extract (Formedium™)
TYM medium	Bacto-tryptone (10 g) Yeast extract (2.5g) NaCl (2.92 g) 1 M $MgCl_2$ (5 ml) dH_2O up to 1 l
Bacterial M9 minimum medium	5x M9 salts (40 ml) 2 mM glucose (2.75 ml) water agar (157.25 ml)
5x M9 salt solution	Na_2HPO_4 x 7 H_2O (6.78 g) KH_2PO_4 (3.0 g) NaCl (0.5 g) NH_4Cl (1.0 g)
Water agar	Bactoagar (3.0 g) dH_2O (175.25 ml)
TfbI resuspending solution	0.1 M KOAc (30 ml) 0.5 M $MnCl_2$ (10 ml) 1 M KCl (10 ml) 1 M $CaCl_2$ (1 ml) Glycerol (15 ml) dH_2O up to 100 ml
Tfb II solution	0.1 M Na-MOPS (2.5 ml) 1 M $CaCl_2$ (1.875 ml) 1 M KCl (0.25 ml) Glycerol (3.75 ml) dH_2O up to 25 ml

Buffers and solutions

0.5 x TBE buffer	45 mM Tris base (Fisher Scientific) 45 mM Boric acid (Fisher Scientific) 1 mM EDTA (Fisher Scientific)
0.5 M EDTA	EDTA (14.61 g), dissolve powder in 80 ml water, adjust pH to 8.0 Molecular grade water up to 100 ml
2x RNA Loading buffer	95% (v/v) formamide 5 mM EDTA 1.9 mM xylene cyanol 1.5 mM bromophenol blue
10 x NEB2 buffer	500 mM NaCl 100 mM Tris-HCl (pH 8.0) 100 mM $MgCl_2$ 10 mM DTT
10% SDS	10 g of sodium dodecylsulfate were dissolved in 100 ml water
20x SSC	3M NaCl (175.3 g) 0.3 M Sodium citrate (88.2 g) Dissolve in 800 ml water Adjust pH at 7.0 Fill up with water to 1l
Northern blot washing buffer	0.2x SSC, 0.1 % SDS
Northern blot stripping buffer	0.1% SDS (10 ml 10% SDS + 990 ml water)
10x TBST-buffer	200 mM Tris 1.5 M NaCl 1% Tween-20
10x PBS buffer	130 mM NaCl 774 mM Na_2HPO_4 226 mM $Na_2H_2PO_4$

Appendix II

II.1 Oligonucleotides used for generating 3' end hY5 RNA mutants

Table 0.1. List of DNA oligonucleotides used for hY5 gene amplification.

In this table the DNA oligonucleotides that were used to amplify the entire hY5 RNA gene (including promoter and terminator) from genomic DNA are listed. The forward primer binds at the promoter and the reverse primer binds to the terminator of the hY5 RNA.

Oligonucleotide	Sequence from 5' to 3'
hY5 forward primer	AATACTAGTGAAGATCCATGGAGGTACATC
hY5 reverse primer	GTAAACGTTGTCTACTACTGTTATTAGTGC

Table 0.2. DNA oligonucleotides used for colony PCR.

This table shows the DNA oligonucleotide sequences that are added to amplify the insert during colony PCR. These primers were used to confirm the successful insertion of the PCR product into the pGEMT easy vector.

Oligonucleotide	**Sequence from 5' to 3'**
pUC M13 forward primer	CGCCAGGGTTTTCCCAGTCACGAC
pUC M13 reverse primer	TCACACAGGAAACAGCTATGAC

Table 0.3. DNA oligonucleotides used for generating the 3' end hY5 RNA mutants.

This table shows the DNA oligonucleotide sequences that were used to introduce the mutations for the generation of 3' end hY5 RNA mutants.

Oligonucleotide	**Sequence from 5' to 3'**
Forward primer	AATACTAGTGAAGATCCATGGAGGTACATC
L1-1 ACGTC	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTGACGTGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L1-2 ACCGC	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTGCGGTGGGGAGACAATGTTAAA TCAACTTAACAATAACCC
L1-3 CACAC	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTGTGTGGGGGAGACAATGTTAAATCAACTTAACAATA ACCC

L1-4 ATGGC	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTGCCATGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L1-5 ACTGC	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTGCAGTGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L1-6 GTCGG	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTCCGACGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L1-7 TCAAA	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTTTGAGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L1-8 TTGAA	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTTCAAGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L1-9 TCGCG	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTCGCGAGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L1-10 TCAAG	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTCTTGAGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L1-CCACT	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTAGTGGGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L1-CCATC	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTGATGGGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L1-CTACG	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTCGTAGGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L1-GACCT	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTAGGTCGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L1-GATCC	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTGGATCGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L1-GAACC	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTGGTTCGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L1-CTACA	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTAGGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L1-TCACA	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGAGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L2-1 CCGTT	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L2-2 ACGTC	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCAACGGTGTGGGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L2-3 CCGTC	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGACGTTGTGGGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L2-4 CTGTC	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGACGGTGTGGGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L2-5 CTGTG	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGACAGTGTGGGGGGAGACAATGTTAAATCAACTTAACAATA ACCC

L2-6 GTCGG	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCCACTGTGTGGGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L2-7 TCAAA	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGAGGTTGTGGGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L2-8 TTGAA	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCTTGTGTGGGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L2-9 TCGCG	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCCCGAGTGTGGGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L2-10 TCAAG	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGAACATGTGGGGGGAGACAATGTTAAATCAACTTAACAATA ACCC
L3-1 GTTTG	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAACGCGGTTGTGGGGGGAGACAATGTTA
L3-2 TGTAG	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CTACAGCGGTTGTGGGGGGAGACAATGTTA
L3-3 GTTCG	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CGAACGCGGTTGTGGGGGGAGACAATGTTA
L3-4 CGTTG	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAACGGCGGTTGTGGGGGGAGACAATGTTA
L3-5 CGTCG	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CGACGGCGGTTGTGGGGGGAGACAATGTTA
L3-6 TCCCA	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT TGGGAGCGGTTGTGGGGGGAGACAATGTTA
L3-7 TAATT	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT AATTAGCGGTTGTGGGGGGAGACAATGTTA
L3-8 GAAAC	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT ATTTCGCGGTTGTGGGGGGAGACAATGTTA
L3-9 ATCTA	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT TAGATGCGGTTGTGGGGGGAGACAATGTTA
L3-10 CGACC	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT GGTCGGCGGTTGTGGGGGGAGACAATGTTA
T49 to A49	AGTTGGTCCGAGTGTTGTGGGTTATTGTTAAGTTGATTTAAC ATTGTCACCCCCCACAACCGCG
T49 to C49	AGTTGGTCCGAGTGTTGTGGGTTATTGTTAAGTTGATTTAAC ATTGTCACCCCCCCCAACCGCG
T49 to G49	AGTTGGTCCGAGTGTTGTGGGTTATTGTTAAGTTGATTTAAC ATTGTCACCCCCCGCAACCGCG
Δ 1 nt	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGGGAGACATGTTAAATCAACTTAACA
Δ 2 nt	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGGGAGACTGTTAAATCAACTTAACA
Δ 3 nt	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGGGAGATGTTAAATCAACTTAACA
Δ 4 nt	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGGGATGTTAAATCAACTTAACA
Δ 5 nt	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGGGATGTTAAATCAACTTAACA
Δ 6 nt	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGGGTGTTAAATCAACTTAACA
Δ 7 nt	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGGTGTTAAATCAACTTAACA
Δ 8 nt	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGTGTTAAATCAACTTAACA
Δ 9 nt	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGTGTTAAATCAACTTAACA

Loop M1	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGCGGAGACAATGTTAAATCAACTTA
Loop M2	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGCGAGACAATGTTAAATCAACTTA
Loop M3	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGAGAGACAATGTTAAATCAACTTA
Loop M4	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGAGGAGACAATGTTAAATCAACTTA
Loop M5	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGGCAGACAATGTTAAATCAACTTA
Loop M6	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGCGAGACAATGTTAAATCAACTTA
Loop M7	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGTGAGACAATGTTAAATCAACTTA
Loop M8	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGAGAGACAATGTTAAATCAACTTA
Loop M9	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGTTATTGTTAAATCAACTTA
Loop M10	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGCCAGACAATGTTAAATC
Loop M11	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGACAGACAATGTTAAATC
Loop M12	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGCAAGACAATGTTAAATC
Loop M13	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGCAAGACAATGTTAAATC
Loop M14	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGGAGACAATGTTAAATCAACTTA
Loop M15	GTAAACGTTGTCTACTACTGTTATTAGTGCAAAACAGCAAGCTAGT CAAGCGCGGTTGTGGGAGACAATGTTAAATCAACTTA

Table 0.4. DNA oligonucleotides used for generating "balancing" DNA.

In this table the DNA oligonucleotides that were used to generate the "balancing DNA" are listed. As the full length and plasmid bias hY5 RNA libraries start with the same 15 nucleotides a so called "balancing DNA" sequence was generated in order to introduce more heterogeneity and more sequence diversity into the sequencing pool so that the different library sets could be sequenced on one lane and be clustered separately.

Oligonucleotide	Sequence
Balancing DNA template	CGATCAGTTGGTCCGAGTGTTGTGGGTTATTGTTAAGTTGATTTAACNNNNNNNNNNNNNNNNNNNNNNATTGTCTCCAGCCACAACCGCTGGAATTCTCGGG
Forward primer	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGACGATCAGTTGGTCCGAGTGTTGTGG
Reverse primer	CAAGCAGAAGACGGCATACGATACGAGATATATTAGTGACTGGAGTTCCATGGCACCCGAGAATTCCAGCGGTTGTGGC
Balancing DNA	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGACGATCAGTTGGTCCGAGTGTTGTGGGTTATTGTTAAGTTGATTTAACNNNNNNNNNNNNNNNNNNNNNNATTGTCTCCAGCCACAACCGCTGGAATTCTCGGGTGCCAAGGAACTCCAGTCACTAATATATCTCGTATGCCGTCTTCTGCTTG
Sequencing primer for full length and plasmid libraries	GTCCGAGTGTTGTGGGTTATTGTTAAGTTGATTTAAC

II.2 Oligonucleotides used for generating 5' end hY5 RNA mutants

Table 0.5. DNA oligonucleotides used to generate the hY5 mutant pool PCR products in strategy 1.

In this table the DNA oligonucleotides that were used in the first strategy to introduce the mutations into the hY5 RNA by PCR are shown. In the first PCR reaction the forward primer and overlap reverse primer for mutant pool 4,5 and 6 were added to generate the first fragment. In a second PCR reaction the mutant pool 4, 5 and 6 forward mutagenesis short primers and the hY5 reverse primer were used. The Ns indicate the random nucleotides that were introduced by PCR. The full length hY5 RNA sequence containing the introduced mutations was amplified in an overlap extension PCR.

Oligonucleotide	Sequence from 5' to 3'
Forward primer	AATACTAGTGAAGATCCATGGAGGTACATC
Overlap reverse primer Pool 4 & 5	CAACACTCGGACCAACTGTGTTATCCT
Overlap reverse primer for Pool 6	CAATAACCCACAACACTCGGACCAACTGTGTT

Pool 4 forward mutagenesis short primer	AGGATAACACAGTTGGTCCGAGTGTTGNNNNNTAT TGTTAAGTTGATTTAACATTGTCTCCCCCCACACC GC
Pool 5 forward mutagenesis short primer	AGTTGGTCCGAGTGTTGTGGGTNNNNNTTAAGTTG ATTTAACATTGTCTCCCCCCACAACCGCG
Pool 6 forward mutagenesis short primer	GTCCGAGTGTTGTGGGTTATTGNNNNNTTGATTTA ACATTGTCTCCCCCCACAACCGCG
Reverse primer	GTAAACGTTGTCTACTACTGTTATTAGTGC

Table 0.6. DNA oligonucleotides used for generating 5' end hY5 plasmid libraries.

This table shows the DNA oligonucleotides that were used to amplify the PCR products during the generation of 5' end hY5 plasmid bias libraries. For each library replicate a different Illumina multiplex sequencing index primer containing a unique 6 nt barcoded (shown the table underlined) in order to identify individual samples after Illumina sequencing.

Oligonucleotide	Sequence 5' to 3'
First PCR reaction forward primer	GTTCAGAGTTCTACAGTCCGACGATCTAACACAGTT GGTCCGAGTGTT
First PCR reaction reverse primer	CCTTGGCACCCGAGAATTCCATTAGTGCAAAACAGC AAGCT
RP-3	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACT GGAGTTCCTTGGCACCCGAGAATTCCA
RP-7	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACT GGAGTTCCTTGGCACCCGAGAATTCCA
RP-10	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACT GGAGTTCCTTGGCACCCGAGAATTCCA
RP-17	CAAGCAGAAGACGGCATACGAGATCTCTACGTGACT GGAGTTCCTTGGCACCCGAGAATTCCA
RP-21	CAAGCAGAAGACGGCATACGAGATCGAAACGTGACT GGAGTTCCTTGGCACCCGAGAATTCCA
RP-23	CAAGCAGAAGACGGCATACGAGATCCACTCGTGACT GGAGTTCCTTGGCACCCGAGAATTCCA
RP-5	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACT GGAGTTCCTTGGCACCCGAGAATTCCA
RP-6	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACT GGAGTTCCTTGGCACCCGAGAATTCCA
RP-8	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACT GGAGTTCCTTGGCACCCGAGAATTCCA

Table 0.7. DNA oligonucleotides used to generate the hY5 mutant pool PCR products in strategy 2.

In this table the DNA oligonucleotides that were used in the second strategy to introduce the mutations into the hY5 RNA by PCR are shown. In the second strategy the full length hY5 RNA sequence had to be amplified using a truncated hY5 RNA sequence in order to avoid the overamplification of the hY5 RNA sequence. In the first PCR reaction to generate the plasmid DNA template 1 the forward primer and overlap reverse primers for pool 4, 5 and 6 were used. For the plasmid DNA template 2 the forward primer for short template and reverse primer for short template were used. The PCR products were cloned into pGEMT easy vector and the resulting plasmid vectors were used as templates to introduce the mutations using the pool 4, 5 & 6 forward mutagenesis primers and the reverse primer. The short template was generated using the forward primer and overlap reverse primers. The full length hY5 RNA containing the mutagenized regions were amplified by overlap extension PCR using the forward primer and reverse primer. The Ns indicate the random nucleotides that were introduced by PCR.

Oligonucleotide	Sequence 5' to 3'
Forward primer	AATACTAGTGAAGATCCATGGAGGTACATC
Overlap reverse primer Pool 4 & 5	CAACACTCGGACCAACTGTGTTATCCT
Overlap reverse primer for Pool 6	CAATAACCCACAACACTCGGACCAACTGTGTT
Forward primer for short template	TTGATTTAACATTGTCTCCCCCCACAACCG
Reverse primer for short template	GTAAACGTTGTCTACTACTGTTATTAGTGC
Pool 4 forward mutagenesis primer	AGGATAACACAGTTGGTCCGAGTGTTGNNNNNTATTG TTAAGTTGATTTAACATTGTCTCCCCCCACACCGC
Pool 5 forward mutagenesis primer	AGTTGGTCCGAGTGTTGTGGGTNNNNNTTAAGTTGAT TTAACATTGTCTCCCCCCACAACCGCG
Pool 6 forward mutagenesis primer	GTCCGAGTGTTGTGGGTTATTGNNNNNTTGATTTAAC ATTGTCTCCCCCCACAACCGCG
Reverse primer	GTAAACGTTGTCTACTACTGTTATTAGTGC

Table 0.8. DNA oligonucleotides used for the final overlap extension PCR.

In the table the DNA oligonucleotides are listed that were used to generate the whole hY5 RNA gene including mutated regions and promoter and terminator sequence.

Oligonucleotide	Sequence 5' to 3'
Forward primer	AATACTAGTGAAGATCCATGGAGGTACATC
Reverse primer	GTAAACGTTGTCTACTACTGTTATTAGTGC

Table 0.9. DNA oligonucleotides used for generating full length cDNA libraries

The table shows the DNA oligonucleotides that were used for reverse transcription and generation of the hY5 full length cDNA libraries. The Illumina multiplex index sequences contain a 6 nt barcode sequence (underlined in the table) that were introduced by PCR for sample identification.

Oligonucleotide	Sequence 5' to 3'
hY5 full length RT primer	AAACAGCAAGCTAGT
First PCR reaction FL forward primer	GTTCAGAGTTCTACAGTCCGACGATCAGTTGGTCCGAGTGTTG
First PCR reaction FL reverse primer	CCTTGGCACCCGAGAATTCCAAACAGCAAGCTAGTCAAGCG
RP5	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RP6	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RP8	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RP9	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RP10	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RP12	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RP1	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RP2	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RP3	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA
RP19	CAAGCAGAAGACGGCATACGAGATTTTCACGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA

Table 0.10. DNA oligonucleotides used for small RNA library construction of 5' end derived YsRNA mutants.

In the table YsRNA enrichment sequences with 5, 10 or 15 nt that were tested for YsRNA enrichment during YsRNA library construction are listed.

Oligonucleotide	Sequence 5' to 3'
RP1 hY5 enrichment 5 nt	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGACGACGATCAGTTG
RP1 hY5 enrichment 10 nt	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGACGACGATCAGTTGGTCCG
RP1 hY5 enrichment 15 nt	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGACGACGATCAGTTGGTCCGAGTGT

Table 0.11. DNA oligonucleotides used for generating individual 5' end derived hY5 RNA mutant sequences.

Oligonucleotide	Sequence from 5' to 3'
Forward primer	AATACTAGTGAAGATCCATGGAGGTACATC
Overlap reverse primer Pool 4 & 5	CAACACTCGGACCAACTGTGTTATCCT
Overlap reverse primer for Pool 6	CAATAACCCACAACACTCGGACCAACTGTGTT
Mutagenesis primer	
Pool 4 CAACC	AGGATAACACAGTTGGTCCGAGTGTTGCAACC TATTGTTAAGTTGATTTAACATTGTCTCCCCC CACAACCGCG
Pool 4 TGGTT	AGGATAACACAGTTGGTCCGAGTGTTGTGGTT TATTGTTAAGTTGATTTAACATTGTCTCCCCC CACAACCGCG
Pool 4 CGGGT	AGGATAACACAGTTGGTCCGAGTGTTGCGGGT TATTGTTAAGTTGATTTAACATTGTCTCCCCC CACAACCGCG
Pool 4 AGGGT	AGGATAACACAGTTGGTCCGAGTGTTGAGGGT TATTGTTAAGTTGATTTAACATTGTCTCCCCC CACAACCGCG
Pool 4 TGAGT	AGGATAACACAGTTGGTCCGAGTGTTGTGAGT TATTGTTAAGTTGATTTAACATTGTCTCCCCC CACAACCGCG
Pool 4 TGGCT	AGGATAACACAGTTGGTCCGAGTGTTGTGGCT TATTGTTAAGTTGATTTAACATTGTCTCCCCC CACAACCGCG
Pool 4 TGGGA	AGGATAACACAGTTGGTCCGAGTGTTGTGGGA TATTGTTAAGTTGATTTAACATTGTCTCCCCC CACAACCGC
Pool 4 TGGGC	AGGATAACACAGTTGGTCCGAGTGTTGTGGGC TATTGTTAAGTTGATTTAACATTGTCTCCCCC CACAACCGC
Pool 4 TGGGG	AGGATAACACAGTTGGTCCGAGTGTTGTGGGG TATTGTTAAGTTGATTTAACATTGTCTCCCCC CACAACCGC
Pool 5 TATTG	AGTTGGTCCGAGTGTTGTGGGTTATTGTTAAG TTGATTTAACATTGTCTCCCCCCACAACCGCG
Pool 5 ATCGC	AGTTGGTCCGAGTGTTGTGGGTATCGCTTAAG TTGATTTAACATTGTCTCCCCCCACAACCGCG
Pool 5 TATGG	AGTTGGTCCGAGTGTTGTGGGTTATGGTTAAG TTGATTTAACATTGTCTCCCCCCACAACCGCG
Pool 5 GTTCG	AGTTGGTCCGAGTGTTGTGGGTGTTCGTTAAG TTGATTTAACATTGTCTCCCCCCACAACCGCG
Pool 5 TGTCG	AGTTGGTCCGAGTGTTGTGGGTTGTCGTTAAG TTGATTTAACATTGTCTCCCCCCACAACCGCG

Pool 5 TATGA	AGTTGGTCCGAGTGTTGTGGGTTATGATTAAG TTGATTTAACATTGTCTCCCCCCACAACCGCG
Pool 5 TCCGC	AGTTGGTCCGAGTGTTGTGGGTTCCGCTTAAG TTGATTTAACATTGTCTCCCCCCACAACCGCG
Pool 5 TTCCG	AGTTGGTCCGAGTGTTGTGGGTTTCCGTTAAG TTGATTTAACATTGTCTCCCCCCACAACCGCG
Pool 5 GGGCG	AGTTGGTCCGAGTGTTGTGGGTGGGCGTTAAG TTGATTTAACATTGTCTCCCCCCACAACCGCG
Pool 5 TCGAC	AGTTGGTCCGAGTGTTGTGGGTTCGACTTAAG TTGATTTAACATTGTCTCCCCCCACAACCGCG
Pool 5 GCAAG	AGTTGGTCCGAGTGTTGTGGGTGCAAGTTAAG TTGATTTAACATTGTCTCCCCCCACAACCGCG
Pool 5 ACGTA	AGTTGGTCCGAGTGTTGTGGGTACGTATTAAG TTGATTTAACATTGTCTCCCCCCACAACCGCG
Pool 5 CACC	AGTTGGTCCGAGTGTTGTGGGCACCTGTTAAG TTGATTTAACATTGTCTCCCCCCACAACCGCG
Pool 5 AAAA	AGTTGGTCCGAGTGTTGTGGGAAAATGTTAAG TTGATTTAACATTGTCTCCCCCCACAACCGCG
Pool 5 ATAT	AGTTGGTCCGAGTGTTGTGGGATATTGTTAAG TTGATTTAACATTGTCTCCCCCCACAACCGCG
Pool 5 AAAT	AGTTGGTCCGAGTGTTGTGGGAAATTGTTAAG TTGATTTAACATTGTCTCCCCCCACAACCGCG
Pool 5 CCCC	AGTTGGTCCGAGTGTTGTGGGCCCCTGTTAAG TTGATTTAACATTGTCTCCCCCCACAACCGCG
Pool 5 CGTT	AGTTGGTCCGAGTGTTGTGGGCGTTTGTTAAG TTGATTTAACATTGTCTCCCCCCACAACCGCG
Pool 5 TTTT	AGTTGGTCCGAGTGTTGTGGGTTTTTGTTAAG TTGATTTAACATTGTCTCCCCCCACAACCGCG
Pool 6 TTAAG	GTCCGAGTGTTGTGGGTTATTGTTAAGTTGAT TTAACATTGTCTCCCCCCACAACCGCG
Pool 6 GGGAG	GTCCGAGTGTTGTGGGTTATTGGGGAGTTGAT TTAACATTGTCTCCCCCCACAACCGCG
Pool 6 GAGGG	GTCCGAGTGTTGTGGGTTATTGGAGGGTTGAT TTAACATTGTCTCCCCCCACAACCGCG
Pool 6 AGGAG	GTCCGAGTGTTGTGGGTTATTGAGGAGTTGAT TTAACATTGTCTCCCCCCACAACCGCG
Pool 6 TGGTT	GTCCGAGTGTTGTGGGTTATTGTGGTTTTGAT TTAACATTGTCTCCCCCCACAACCGCG
Pool 6 TGGTG	GTCCGAGTGTTGTGGGTTATTGTGGTGTTGAT TTAACATTGTCTCCCCCCACAACCGCG
Pool 6 GGGGA	GTCCGAGTGTTGTGGGTTATTGGGGGATTGAT TTAACATTGTCTCCCCCCACAACCGCG
Pool 6 GGGTG	GTCCGAGTGTTGTGGGTTATTGGGGTGTTGAT TTAACATTGTCTCCCCCCACAACCGCG

II.3 Oligonucleotides used for Northern blotting

Table 0.12. List of DNA oligonucleotides used for Northern blotting.

In this table DNA oligonucleotides used as Northern blot probes are shown. The DNA oligonucleotide sequences were designed to correspond to the antisense sequence of the small RNA sequences of interest.

Northern blot probes	**Sequence from 5' to 3'**
hY5 3' end wild type probe	AGCTAGTCAAGCGCGGTTGTGGGGG
hY5 3' end probe for detection of L1/2 mutants	AGCAAGCTAGTCAAGCGCGGT
LNA 12 mer for detection of L3 mutants	A*CAG*CAA*GCT*AG
U6 snRNA loading control	GCTAATCTTCTCTGTATCGTTCC
hY5 5' end wild type probe	TAACCCACAACACTCGGACCAACT
hY5 5' end probe for detection of L4/5/6 mutants	CAACACTCGGACCAACT
hY1 3' end	AGACTAGTCAAGTGCAGTAGTGAGAA
hY1 5' end	TAACCCACAACACTCGGACCAACT
hY3 3' end	TAGTCAAGTGAAGCAGTGGGAG
hY3 5' end	AACACCACTGCACTCGGACCAGCC
tRNA-His	CAGAGTACTAACCACTATACGATCACGGCC
tRNA-Pro	CCGAGAATCATACCCCTAGACCAACGAGCC

Appendix III

Table 0.13. Sequencing information of 3' end hY5 small RNA cDNA libraries.

The table gives an overview about the number of total sequencing reads, reads mapping to the genome and reads mapping to hY5 RNA mutant sequences including the percentages of genome mapping reads and hY5 RNA mapping reads for each library replicate. The libraries 1, 2 and 3 are shown in the table as L1, L2 and L3 including their three biological replicates. Total RNA was extracted and YsRNA cDNA libraries were constructed using a modified Illumina protocol and sent for NGS. After adapter removal the sequencing reads were mapped to the genome and all possible 1024 combinations of hY5 mutant sequences.

	Total reads	Genome mapping reads	Genome mapping reads [%]	Y5 mapping reads	Y5 mapping reads [%]
L1-1	3095015	2161090	69.82	56653	1.83
L1-2	3310825	2284251	68.99	70791	2.14
L1-3	2353355	1587079	67.44	44746	1.90
L2-1	5562010	4162303	74.83	95249	1.71
L2-2	2701036	1902208	70.43	107289	3.97
L2-3	2824250	2017230	71.43	105952	3.75
L3-1	1444418	848693	58.76	145191	10.05
L3-2	3787865	2174265	57.40	495631	13.08
L3-3	1153145	677186	58.73	92146	7.99

Appendix IV

Sequencing information of wild type and Ro60 $^{-/-}$ mES cDNA Libraries

Table 0.14. General sequencing information of wild type and Ro60 $^{-/-}$ mES sRNA cDNA libraries either treated or untreated with poly (I:C).

The table gives an overview about the number of total sequencing reads, reads mapping to the genome and reads mapping to hY5 RNA mutant sequences including the percentages of genome mapping reads and hY5 RNA mapping reads for each library replicate. The libraries were generated in three biological replicates. Total RNA was extracted and YsRNA cDNA libraries were constructed using a modified Illumina protocol and sent for NGS. After adapter removal the sequencing reads were mapped to the mouse genome.

	MES WT 1	MES WT 2	MES WT 3	MES WT poly (I:C) 1	MES WT poly (I:C) 2	MES WT poly (I:C) 3
Trimmed	10029505	9004771	8834580	11909230	11677219	14688848
Mapped total	8417398	7185996	7496927	9591872	8397299	12192779
Mapped unique	357069	328363	229033	547665	469518	393481
Mapping %	83.93	79.80	84.86	80.54	71.91	83.01

	Ro60 KO 1	Ro60 KO 2	Ro60 KO 3	Ro60 KO poly(I:C) 1	Ro60 KO poly (I:C) 2	Ro60 KO poly (I:C) 3
Trimmed	12521810	11006607	7809998	16740090	12969950	9942964
Mapped total	10328068	8013893	6457272	13185076	8936481	8496502
Mapped unique	433415	404450	320651	725926	512447	213763
Mapping %	82.48	72.81	82.68	78.76	68.90	85.45

Table 0.15. Differential gene expression analysis of mY1 and mY3 in wild type and Ro60 $^{-/-}$ cells untreated or treated with poly (I:C) using DESeq2.

Locus	baseMean	log2FoldChange	pvalue	padj	Gene ID
6/47787768-47788203	13883.2	6.63395	3.32E-70	2.48E-66	Rny1
6/47781599-47781778	2158.05	2.85508	1.17E-13	1.46E-10	Rny3

Table 0.16. Differential gene expression analysis of SNORD 118, miR 296, miR 146a and miR 708 in wild type and Ro60 $^{-/-}$ cells untreated or treated with poly (I:C) using DEseq2.

Locus	Base Mean	log2 Fold Change	p-value	Sequence	Gene
18/71961786-71961805	67.543	1.343	0.00212	TCTCCAATCATCATTTTCT	SNORD118
12/49279106-49279123	28.712	1.491	0.0021	GAGGGTTGGGTGGAGGCT	miR296
11/43374398-43374457	299.464	1.285	0.00798	TAACCCATGGAATTCAGTTCTCA	miR146a
7/96249426-96249641	3814.868	0.8913	0.00967	AAGGAGCTTACAATCTAGCTGGG	miR708